Improved biological processes for the production of aryl sulfates

Jendresen, Christian Bille; Nielsen, Alex Toftgaard

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 August 2017 (31.08.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/144671 A1

(51) International Patent Classification:
C12P 11/00 (2006.01) *C12P 5/00* (2006.01)
C12N 9/10 (2006.01)

(21) International Application Number:
PCT/EP2017/054346

(22) International Filing Date:
24 February 2017 (24.02.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16157231.8 24 February 2016 (24.02.2016) EP

(71) Applicant: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby (DK).

(72) Inventors: **JENDRESEN, Christian Bille**; Rovsingsgade 15, 3. th., 2100 Copenhagen Ø (DK). **NIELSEN, Alex Toftgaard**; Pennehave 3E, 2960 Rungsted Kyst (DK).

(74) Agent: **ZACCO DENMARK A/S**; Arne Jacobsens Allé 15, 2300 Copenhagen S (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2017/144671 A1

(54) Title: IMPROVED BIOLOGICAL PROCESSES FOR THE PRODUCTION OF ARYL SULFATES

(57) Abstract: The present invention generally relates to the field of biotechnology as it applies to the production of aryl sulfates using recombinant host cells. More particularly, the present invention pertains to recombinant host cells comprising (e.g., expressing) a polypeptide having aryl sulfotransferase activity, wherein said recombinant host cells have been modified to have an increased uptake of sulfate compared to identical host cells that does not carry said modification. Further provided are processes for the production of aryl sulfates, such as zosteric acid, employing such recombinant host cells.

Improved biological processes for the production of aryl sulfates

Technical field of the invention

The present invention generally relates to the field of biotechnology as it applies to the production of aryl sulfates using recombinant host cells. More particularly, the present invention pertains to recombinant host cells comprising (e.g., expressing) a polypeptide having aryl sulfotransferase activity, the use of such recombinant host cells in the production of aryl sulfates, and processes for the production of aryl sulfates, such as zosteric acid, employing such recombinant host cells.

Background of the invention

A range of phenolic compounds are of great interest to the biotech industry since they are building blocks for polymeric compounds. Examples of such phenolic compounds include p-coumaric acid (pHCA) or other hydroxycinnamic acids which form the basis for many secondary metabolites including flavonoids and stilbenes. However, many of these phenolic compounds are toxic to producing organisms, and thus limit the productivity during fermentation. Hence, there is a need for large scale production processes, and especially for biological large scale production processes allowing improved productivity.

Moreover, a range of phenolic compounds, and especially those used as drugs or food additives such as resveratrol or vanillin, show poor solubility in water which makes it difficult for these compounds to be uptaken by the body. Hence, there is also a need for providing such phenolic compounds in a form which improves the solubility, and hence bioavailablility, preferably by using biological large scale production processes.

Summary of the invention

The object of the present invention is to provide a method for large scale production of aryl sulfates. Furthermore, it is an object to provide a biological process for the large scale production of phenols. The inventors have developed a biological process that solves both objects.

The present invention thus provides in a first aspect a process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound, such as p-coumaric acid, with a first recombinant host cell; wherein the first recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having an aryl sulfotransferase activity; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having an aryl sulfotransferase activity; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having an aryl sulfotransferase activity.

Particularly, the present invention provides a process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

The present invention provides in a further aspect a recombinant host cell comprising (e.g. expressing) a heterologous polypeptide having an aryl sulfotransferase activity.

Particularly, the present invention provides a recombinant host cell comprising (e.g. expressing) a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

The present invention can be further summarized by the following items:

1. A process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

2. The process according to item 1, further comprising:

(ii) culturing the first recombinant host cell under suitable conditions for the production of the corresponding sulfated phenolic compound; and

(iii) optionally, recovering said sulfated phenolic compound.

3. The process according to item 1 or 2, wherein the heterologous polypeptide having an aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme.

4. The process according to any one of items 1-3, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

5. The process according to any one of items 1-4, wherein the heterologous polypeptide is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

6. The process according to any one of items 1-5, wherein the first recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide.

7. The process according to item 6, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA

molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide.

8. The process according to item 6 or 7, wherein the exogenous nucleic acid molecule is a vector.

9. The process according to item 6 or 7, wherein the exogenous nucleic acid molecule is stabily integrated into the genome of said first recombinant host cell.

10. The process according to any one of items 1-9, wherein said first recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

11. The process according to item 10, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter.

12. The process according to item 11, wherein the increase in the number of copies of the gene or genes is achieved by introducing into said first recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

13. The process according to any one of items 1-10, wherein said first recombinant host cell comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding a sulfate transporter.

14. The process according to item 12, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the one or more nucleotide sequences encoding said sulfate transporter.

15. The process according to any one of items 12-14, wherein the exogenous nucleic acid molecule is a vector.

16. The process according to any one of items 12-14, wherein the exogenous nucleic acid molecule is stably integrated into the genome of said first recombinant host cell.

17. The process according to item 10, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

18. The process according to any one of item 10, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

19. The process according to any one of items 10-18, wherein the sulfate transporter is a bacterial sulfate transporter.

20. The process according to any one of items 10-19, wherein the sulfate transporter is a selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

21. The process according to any one of items 10-20, wherein the sulfate transporter is a member of the CysZ family.

22. The process according to any one of items 10-21, wherein the sulfate transporter is a CysZ protein.

23. The process according to any one of items 10-22, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

2a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

2b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

2c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

24. The process according to any one of items 10-22, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

3a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15;

3b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15; or

3c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

25. The process according to any one of items 10-20, wherein the sulfate transporter is a sulfate-transporting ATPase.

26. The process according to any one of items 10-20, wherein the sulfate transporter is a member of the SulT (cysPTWA) family.

27. The process according to any one of items 10-20, wherein the sulfate transporter comprises a first membrane subunit (CysT), a second membrane subunit (CysW), an ATP binding subunit (CysA) and a periplasmic binding protein (CysP or Sbp).

28. The process according to any one of items 25 to 27, where the sulfate transporter is encoded by an operon comprising a nucleotide sequence encoding a first membrane subunit (CysT), a nucleotide sequence encoding a second membrane subunit (CysW), a nucleotide sequence encoding an ATP binding subunit (CysA) and a nucleotide sequence encoding a periplasmic binding protein (CysP or Sbp).

29. The process according to item 27 or 28, wherein the first membrane subunit is a polypeptide selected from the group consisting of:

4a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16;

4b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16; or

4c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

30. The process according to any one of items 27 to 29, wherein the second membrane subunit is a polypeptide selected from the group consisting of:

5a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17;

5b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17; or

5c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

31. The process according to any one of items 27 to 30, wherein the ATP binding subunit is a polypeptide selected from the group consisting of:

6a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18;

6b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18; or

6c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

32. The process according to any one of items 27 to 31, wherein the periplasmic binding protein is CysP.

33. The process according to any one of items 27 to 31, wherein the periplasmic binding protein is a polypeptide selected from the group consisting of:

7a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19;

7b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least

about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19; or

7c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

34. The process according to any one of items 27 to 31, wherein the periplasmic binding protein is Sbp.

35. The process according to any one of items 27 to 31, wherein the periplasmic protein is a polypeptide selected from the group consisting of:

8a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20;

8b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20; or

8c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

36. The process according to any one of items 10-20, wherein the sulfate transporter is a member of the SulP family.

37. The process according to any one of items 10-20, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

9a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26;

9b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26; or

9c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

38. The process according to any one of items 10-20, wherein the sulfate transporter is a CysP transporter belonging to the phosphate inorganic transporter (PiT) family.

39. The process according to any one of items 10-20, wherein the sulfate transporter is selected from the group consisting of:

10a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27;

10b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; or

10c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

40. The process according to any one of items 1-39, wherein the first recombinant host cell has been further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

41. The process according to item 40, wherein the ATP sulfurylase is encoded by the genes cysD and cysN.

42. The process according to item 40 or 41, wherein the ATP sulfurylase comprises a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least

95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29.

43. The process according to any one of items 1-42, wherein said first recombinant host cell has been further modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

44. The process according to item 43, wherein the APS kinase is encoded by the gene cysC.

45. The process according to item 43 or 44, wherein the APS kinase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32.

46. The process according to any one of items 1-45, wherein said first recombinant host cell has been further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

47. The process according to item 46, wherein said PAP phosphatase is encoded by the gene cycQ.

48. The process according to item 46 or 47, wherein the PAP phosphatase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37.

49. The process according to any one of items 40-48, wherein the increase in protein expression is achieved by increasing the number of copies of the encoding gene or genes.

50. The process according to item 49 wherein the increase in the number of copies of the gene or genes is achieved by introducing into said first recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or

genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

51. The process according to any one of items 40-50, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

52. The process according to any one of items 40-51, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes.

53. The process according to any one of items 1-52, wherein said first recombinant host cell further comprises a heterologous polypeptide having a tyrosine ammonia lyase activity.

54. The process according to any one of items 1-53, wherein in step (i'), (i'') or (i''') the medium is further contacted with a second recombinant host cell comprising a heterologous polypeptide having a tyrosine ammonia lyase activity.

55. The process according to item 53 or 54, wherein the heterologous polypeptide having a tyrosine ammonia lyase activity is selected from the group consisting of:

11a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40);

11b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40); or

11c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

56. The process according to any one of items 53 to 55, wherein the first and/or second recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide having a tyrosine ammonia lyase activity.

57. The process according to item 56, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide.

58. The process according to item 56 or 57, wherein the exogenous nucleic acid molecule is a vector.

59. The process according to item 56 or 57, wherein the exogenous nucleic acid is stably integrated into the genome of the first and/or second recombinant host cell.

60. The process according to any one of items 1 to 59, wherein the first recombinant host cell and the second recombinant host cell are independently selected from the group consisting of bacteria, yeasts, fungi, algae and plant cells.

61. The process according to any one of items 1 to 60, wherein the first recombinant host cell is a bacterium.

62. The process according to item 61, wherein the bacterium is a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus,* or *Yersinia*.

63. The process according to item 61, wherein the bacterium is a bacterium of the genus *Bacillus*.

64. The process according to item 63, wherein the bacterium is *Bacillus subtilis*.

65. The process according to item 61, wherein the bacterium is a bacterium of the genus *Lactococcus.*

66. The process according to item 65, wherein the bacterium is *Lactococcus lactis.*

67. The process according to item 61, wherein the bacterium is a bacterium of the genus *Pseudomonas*.

68. The process according to item 67, wherein the bacterium is *Pseudomonas putida*.

69. The process according to item 61, wherein the bacterium is a bacterium of the genus *Corynebacterium*.

70. The process according to item 69, wherein the bacterium is *Corynebacterium glutamicum*.

71. The process according to item 61, wherein the bacterium is a bacterium of the genus *Escherichia*.

72. The process according to item 71, wherein the bacterium is *Escherichia coli*.

73. The process according to any one of item 1-60, wherein the first recombinant host cell is a yeast.

74. The process according to item 73, wherein the yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

75. The process according to any one of items 1-60, wherein the first recombinant host cell is a fungus.

76. The process according to item 75, wherein the fungus is a fungus of the genus *Aspergillus*.

77. The process according to any one of items 1-60, wherein the first recombinant host cell is an algae cell.

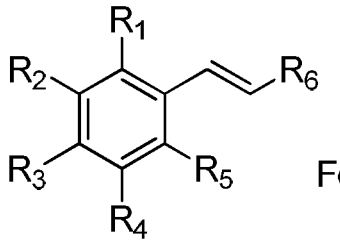
78. The process according to item 77, wherein the algae cells is an algae cell of the genus *Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

79. The process according to any one of items 1-60, wherein the first recombinant host cell is a plant cell.

80. The process according to item 79, wherein the plant cell is selected from the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar

beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

81. The process according to any one of items 1-80, wherein the phenolic compound is represented by the general formula (I):



Formula (I)

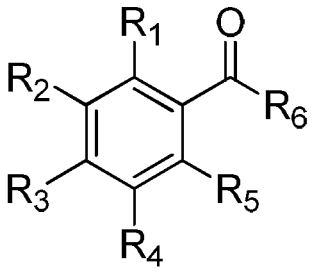
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

82. A process according to any one of the items 1-81, wherein the phenolic compound is represented by the general formula (II):



Formula (II)

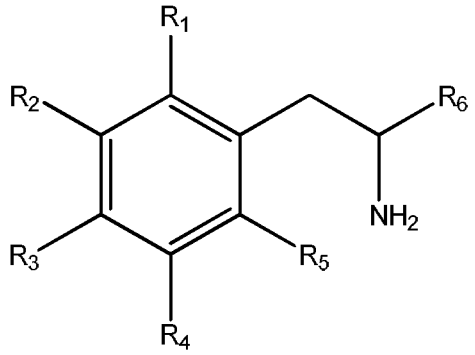
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

83. The process according to any one of items 1-82, wherein the precursor of a phenolic compound in step (i''') is a compound of the general Formula (p-I):



Formula (p-I);

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0

to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

84. The process according to any one of items 81-83, wherein R_6 is $-COOR_7$, wherein R_7 is selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl.

85. The process according to item 84, wherein R_7 is hydrogen.

86. The process according to any one of items 81-85, wherein R_3 is hydroxyl (-OH).

87. The process according to any one of items 81-86, wherein each of R_1, R_2, R_4 and R_5 is hydrogen.

88. The process according to any one of items 81-86, wherein R_4 is hydroxyl (-OH).

89. The process according to item 88, wherein each of R_1, R_2, and R_5 is hydrogen.

90. The process according to any one of items 81-83, wherein each of R_1, R_3 and R_5 is hydrogen, each of R_2 and R_4 is hydroxyl (-OH), and R_6 is p-hydroxyphenyl.

91. A recombinant host cell comprising a heterologous polypeptide having aryl sulfotransferase activity, such as a polypeptide selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted;

wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

92. The recombinant host cell according to item 91, wherein the heterologous polypeptide having an aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme.

93. The recombinant host cell according to item 91, wherein the heterologous polypeptide is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1 ;

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

94. The recombinant host cell according to item 91 to 93, wherein the polypeptide according to 1b) or 1c) has aryl sulfotransferase activity.

95. The recombinant host cells according to any one of items 91-94, the host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide having aryl sulfotransferase activity.

96. The recombinant host cell according to item 95, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide having aryl sulfotransferase activity.

97. The recombinant host cell according to item 96, wherein the exogenous nucleic acid molecule further comprises at least one regulatory element selected from a 5' untranslated region (5'UTR) and 3' untranslated region (3' UTR).

98. The recombinant host cell according to any one of items 95-97, wherein the exogenous nucleic acid is a vector.

99. The recombinant host cell according to any one of items 95-97, wherein the exogenous nucleic acid is stably integrated into the genome of the host cell.

100. The recombinant host cell according to any one of items 91-99, wherein said recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

101. The recombinant host cell according to item 100, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter.

102. The recombinant host cell according to item 101, wherein the increase in the number of copies of the gene or genes is achieved by introducing into said recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

103. The recombinant host cell according to any one of items 91-100, wherein said recombinant host cell comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding a sulfate transporter.

104. The recombinant host cell according to item 102, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the one or more nucleotide sequences encoding said sulfate transporter.

105. The recombinant host cell according to any one of items 102-104, wherein the exogenous nucleic acid molecule is a vector.

106. The recombinant host cell according to any one of items 102-104, wherein the exogenous nucleic acid molecule is stably integrated into the genome of said recombinant host cell.

107. The recombinant host cell according to item 100, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

108. The recombinant host cell according to any one of item 100, wherein the increase in protein expression is achieved by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

109. The recombinant host cell according to any one of items 100-108, wherein the sulfate transporter is a bacterial sulfate transporter.

110. The recombinant host cell according to any one of items 100-109, wherein the sulfate transporter is a selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

111. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is a member of the CysZ family.

112. The recombinant host cell according to any one of items 100-111, wherein the sulfate transporter is a CysZ protein.

113. The recombinant host cell according to any one of items 100-112, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

2a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

2b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

2c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

114. The recombinant host cell according to any one of items 100-112, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

3a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15;

3b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15; or

3c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

115. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is a sulfate-transporting ATPase.

116. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is a member of the SulT (cysPTWA) family.

117. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter comprises a first membrane subunit (CysT), a second membrane subunit (CysW), an ATP binding subunit (CysA) and a periplasmic binding protein (CysP or Sbp).

118. The recombinant host cell according to any one of items 115 to 117, where the sulfate transporter is encoded by an operon comprising a nucleotide sequence encoding a first membrane subunit (CysT), a nucleotide sequence encoding a second membrane subunit (CysW), a nucleotide sequence encoding an ATP binding subunit (CysA) and a nucleotide sequence encoding a periplasmic binding protein (CysP or Sbp).

119. The recombinant host cell according to item 117 or 118, wherein the first membrane subunit is a polypeptide selected from the group consisting of:

4a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16;

4b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16; or

4c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

120. The recombinant host cell according to any one of items 117 to 119, wherein the second membrane subunit is a polypeptide selected from the group consisting of:

5a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17;

5b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17; or

5c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

121. The recombinant host cell according to any one of items 117 to 120, wherein the ATP binding subunit is a polypeptide selected from the group consisting of:

6a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18;

6b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18 or 23; or

6c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

122. The recombinant host cell according to any one of items 117 to 121, wherein the periplasmic binding protein is CysP.

123. The recombinant host cell according to any one of items 117 to 121, wherein the periplasmic binding protein is a polypeptide selected from the group consisting of:

7a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19;

7b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19; or

7c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

124. The recombinant host cell according to any one of items 117 to 121, wherein the periplasmic binding protein is Sbp.

125. The recombinant host cell according to any one of items 117 to 121, wherein the periplasmic protein is a polypeptide selected from the group consisting of:

8a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20;

8b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20; or

8c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

126. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is a member of the SulP family.

127. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is a polypeptide selected from the group consisting of:

9a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26;

9b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26; or

9c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

128. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is a CysP transporter belonging to the phosphate inorganic transporter (PiT) family.

129. The recombinant host cell according to any one of items 100-110, wherein the sulfate transporter is selected from the group consisting of:

10a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27;

10b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; or

10c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

130. The recombinant host cell according to any one of items 91-129, wherein the recombinant host cell has further been modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification.

131. The recombinant host cell according to item 130, wherein said ATP sulfurylase is encoded by the genes cysD and cysN.

132. The recombinant host cell according to item 130 or 131, wherein the ATP sulfurylase comprises a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29.

133. The recombinant host cell according to any one of items 91-132, wherein the recombinant host cell has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

134. The recombinant host cell according to item 133, wherein said APS kinase is encoded by the gene cysC.

135. The process according to item 133 or 134, wherein the APS kinase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32.

136. The recombinant host cell according to any one of items 91-135, wherein the recombinant host cell has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

137. The recombinant host cell according to item 136, wherein said PAP phosphatase is encoded by the gene cycQ.

138. The process according to item 136 or 137, wherein the PAP phosphatase is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37.

139. The recombinant host cell according to any one of items 91-138, wherein the increase in gene expression has been achieved by an increased number of copies of the gene or genes.

140. The recombinant host cell according to item 139, wherein the increase in the number of copies of the gene or genes is achieved by having introduced into said first recombinant host cell one or more exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule..

141. The recombinant host cell according to any one of item 130-138, wherein the increase in protein expression is achieved by modifying the ribosome binding site.

142. The recombinant host cell according to any one of items 130-138, wherein the increase in gene expression has been achieved by increasing the strength of the promoter(s) operably linked to the gene or genes.

143. The recombinant host cell according to any one of items 91-142, further comprising a heterologous polypeptide having tyrosine ammonia lyase activity, such as a polypeptide selected from the group consisting of:

11a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40);

11b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40); or

11c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

144. The recombinant host cell according to item 143, wherein the heterologous polypeptide according to 10b) or 10c) has tyrosine ammonia lyase activity.

145. The recombinant host cell according to item 143 or 144, wherein the recombinant host cell comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding said heterologous polypeptide having tyrosine ammonia lyase activity.

146. The recombinant host cell according to item 145, wherein the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said heterologous polypeptide having tyrosine ammonia lyase activity.

147. The recombinant host cell according to item 145 or 146, wherein the exogenous nucleic acid molecule is a vector.

148. The recombinant host cell according to item 145 or 146, wherein the exogenous nucleic acid is stably integrated into the genome of the recombinant host cell.

149. The recombinant host cell according to any one of items 91-148, wherein the recombinant host cell is selected from the group consisting of bacteria, yeasts, fungi, algae and plant cells.

150. The recombinant host cell according to any one of items 910-149, wherein the recombinant host cell is a bacterium.

151. The recombinant host cell according to item 150, wherein the bacterium is a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus,* or *Yersinia*.

152. The recombinant host cell according to item 150, wherein the bacterium is a bacterium of the genus *Bacillus*.

153. The recombinant host cell according to item 152, wherein the bacterium is *Bacillus subtilis*.

154. The recombinant host cell according to item 150, wherein the bacterium is a bacterium of the genus *Lactococcus*.

155. The recombinant host cell according to item 154, wherein the bacterium is *Lactococcus lactis*.

156. The recombinant host cell according to item 150, wherein the bacterium is a bacterium of the genus *Pseudomonas*.

157. The recombinant host cell according to item 156, wherein the bacterium is *Pseudomonas putida*.

158. The recombinant host cell according to item 150, wherein the bacterium is a bacterium of the genus *Corynebacterium*.

159. The recombinant host cell according to item 158, wherein the bacterium is *Corynebacterium glutamicum*.

160. The recombinant host cell according to item 150, wherein the bacterium is a bacterium of the genus *Escherichia*.

161. The recombinant host cell according to item 160, wherein the bacterium is *Escherichia coli*.

162. The recombinant host cell according to any one of items 91-149, wherein the recombinant host cell is a yeast.

162. The recombinant host cell according to item 162, wherein the yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

163. The recombinant host cell according to any one of items 91-149, wherein the recombinant host cell is a fungus.

164. The recombinant host cell according to item 163, wherein the fungus is a fungus of the genus *Aspergillus*.

165. The recombinant host cell according to any one of items 91-149, wherein the recombinant host cell is an algae cell.

166. The recombinant host cell according to item 165, wherein the algae cells is an algae cell of the genus *Haematococcus, Phaedactylum, Volvox* or *Dunaliella*.

167. The recombinant host cell according to any one of items 91-149, wherein the recombinant host cell is a plant cell.

168. The recombinant host cell according to item 167, wherein the plant cell is selected from the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

169. The recombinant host cell according to any one of items 91-168, which is employed as first recombinant host cell in the process according to any one of items 1-90.

170. Use of the recombinant host cell according to any one of items 91-169 in the production of a sulfated phenolic compound, e.g., in the production of zosteric acid.

171. Use according to item 170, wherein the sulfated phenolic compound is derived from a phenolic compound of general formula (I) or (II) as defined herein.

Brief description of the drawings

Figure 1: Map of plasmid for expression of SULT1A1 from *Rattus norvegicus* in *Escherichia coli*

Figure 2: Map of plasmid for over-expression of cysDNC in *E. coli*.

Figure 3: Map of plasmid for over-expression of cysDNCQ in *E. coli*.

Figure 4: Toxicity of unsulfated or sulfated products

Figure 5: Map of plasmid for over-expression of cysZ in *E. coli*

Figure 6: Map of plasmid for over-expression of cysPTWA in *E. coli*

Figure 7: Concentrations of zosteric acid in culture media with *E. coli* over-expressing SULT1A1 from *Rattus norvegicus*, either alone or in combination with over-expressing cysDNCQ and cysZ or cysPTWA

Detailed description of the invention

Unless specifically defined herein, all technical and scientific terms used have the same meaning as commonly understood by a skilled artisan in the fields of biochemistry, genetics, and molecular biology.

All methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, with suitable methods and materials being described herein. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will prevail. Further, the materials, methods,

and examples are illustrative only and are not intended to be limiting, unless otherwise specified.

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. See, for example, Current Protocols in Molecular Biology (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA); Molecular Cloning: A Laboratory Manual, Third Edition, (Sambrook et al, 2001, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No. 4,683,195; Nucleic Acid Hybridization (B. D. Harries & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In ENZYMOLOGY (J. Abelson and M. Simon, eds.-in-chief, Academic Press, Inc., New York), specifically, Vols.154 and 155 (Wu et al. eds.) and Vol. 185, "Gene Expression Technology" (D. Goeddel, ed.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); and Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

Recombinant host cells of the invention

In one aspect, the present invention provides a recombinant host cell comprising (e.g., expressing) a heterologous polypeptide having an aryl sulfotransferase activity. More particularly, the present invention provides a recombinant host cell comprising (e.g., expressing) a heterologous polypeptide having an aryl sulfotransferase activity, wherein the recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

Generally, the polypeptide having an aryl sulfotransferase activity employed according to the invention will be heterologous to the host cells, which means that said polypeptide is

normally not found in or made (i.e. expressed) by the host cells, but derived from a different species. Aryl sulfotransferases (EC:2.8.2.1) are a well-defined class of enyzmes catalyzing the transfer of a sulfate group from a donor molecule to an aryl acceptor molecule. This makes them particularly suitable for the sulfation of phenolic compounds such as p-coumaric acid and derivatives thereof (e.g., caffeic acid, ferulic acid or sinapic acid), or resveratrol.

The polypeptide having aryl sulfotransferase activity may be a sulfotransferase 1A1 enzyme, a sulfotransferase 1A2 enzyme, a sulfotransferase 1A3 enzyme, a sulfotransferase 1B1 enzyme, a sulfotransferase 1C1 enzyme, a sulfotransferase 1C2 enzyme, a sulfotransferase 1C4 enzyme, or a sulfotransferase 1E1 enzyme.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1A1 enzyme. According to certain other embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1A2 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1B1 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C1 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C2 enzyme. According to certain embodiments, the the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1C4 enzyme. According to other certain embodiments, the polypeptide having aryl sulfotransferase activity is a sulfotransferase 1E1 enzyme (estrogen sulfotransferase), such as the sulfotransferase 1E1 from *Gallus gallus domesticus*.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a mammalian aryl sulfotransferase, such as a mammalian sulfotransferase 1A1 enzyme.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is an aryl sulfotransferase from *Rattus norvegicus* or a variant thereof. Such variant may have at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence of the aryl sulfotransferase from *Rattus norvegicus*. Such variant may also have an amino acid sequence of the sulfotransferase from *Rattus norvegicus*, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1

to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference aryl sulfotransferase. The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity may be a polypeptide selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 99, 101, 103 or 105 (e.g., SEQ ID NO: 1);

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 99, 101, 103 or 105 (e.g., SEQ ID NO: 1); or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 99, 101, 103 or 105 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity may be a polypeptide selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1);

1b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1); or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the polypeptide having aryl sulfotransferase activity is a polypeptide according to 1a). Accordingly, the polypeptide having aryl sulfotransferase activity may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to particular embodiments, the polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 1. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 2. According to yet other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 3. According to yet other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 4. According to yet other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 5. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 6. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 7. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 8. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 9. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 10. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 11. According other particular embodiments, a polypeptide according to a) comprises an amino acid sequence set forth in SEQ ID NO: 12. According other particular embodiments, a polypeptide according to 1a) comprises an amino acid sequence set forth in SEQ ID NO: 13.

According to other certain embodiments, the polypeptide having aryl sulfotransferase activity is a polypeptide according to 1b). Accordingly, a polypeptide having aryl sulfotransferase activity may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%,

at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1). According to other particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence

identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1. According to other more particular embodiments, a polypeptide according to 1b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 1.

Preferably, a polypeptide according to 1b) has aryl sulfotransferase activity. More preferably, a polypeptide according to 1b) has a aryl sulfotransferase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to certain embodiment, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 4. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 7. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide

comprising the amino acid sequence set forth in SEQ ID NO: 9. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 10. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 11. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 13. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 99. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 101. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 103. According to certain other embodiments, a polypeptide according to 1b) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 105.

With “similar” aryl sulfotransferase activity, it is meant that the polypeptide according to 1b) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the aryl sulfotransferase activity of the reference polypeptide (e.g., SEQ ID NO: 1).

The aryl sulfotransferase activity may for instance be determined in accordance to the following method: Aryl sulfotransferase activity may be determined by the reaction of radioactively sulfur labelled PAPS, [^{35}S]PAPS, with the substrate in presence of the polypeptide of interest. This is described previously, for example by Hattori *et al* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). The reaction takes place in a buffer such as 250 µL

50 mM sodium phosphate pH 6.8 with 1 µM [^{35}S]PAPS (3.7kBq) with 100 µM accepting compound for a period of 30 min at 30°C. The reaction is stopped by addition of 100 µL of a 1:1 mixture of 0.1 M barium acetate and barium hydroxide. 50 µL of 0.1 M zinc sulfate is added, followed by centrifugation at 1,200 × g for 5 min. 300 µL of the supernatant is then transferred to a new container and 50 µL of an equal volume of 0.1 M barium hydroxide and 0.1 M zinc sulfate is added. The mixture is then centrifuged at 13,000 × g for 5 min, and 300-µL aliquots of the supernatant are mixed with 2.5 mL of Cleasol I (Nacalai Tesque, Kyoto, Japan). The radioactivity is then measured by scintillation.

Alternatively, the activity of a sulfotransferase may be detected by direct measurement of the product by analytical methods such as high performance liquid chromatography (HPLC) or liquid chromatography in combination with mass spectrometry (LC-MS).

According to other certain embodiments, the polypeptide having aryl sulfotransferase activity is a polypeptide according to 1c). Accordingly, the polypeptide having aryl sulfotransferase activity may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein 1 or more, such as 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, 100 or more, 110 or more, 120 or more, 130 or more, 140 or more, or 150 or more, amino acid residues are substituted, deleted, and/or inserted. According to particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about

15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1), wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 150, such as about 1 to about 140, about 1 to about 130, about 1 to about 120, about 1 to about 110, about 1 to about 100, about 1 to about 90, about 1 to about 80, about 1 to about 70, about 1 to about 60, about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 1c) comprises an amino acid sequence set forth in SEQ ID NO: 1, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (e.g., SEQ ID NO: 1). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 1c) has aryl sulfotransferase activity. More preferably, a polypeptide according to 1c) has a aryl sulfotransferase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13 (e.g., SEQ ID NO: 1).

According to certain embodiment, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 1. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 2. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 3. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 4. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 5. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 6. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 7. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 8. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 9. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 10.

According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 11. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 12. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 13. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 99. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 101. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 103. According to certain other embodiments, a polypeptide according to 1c) has aryl sulfotransferase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 105.

With "similar" aryl sulfotransferase activity it is meant that the polypeptide according to 1c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the aryl sulfotransferase activity of the reference polypeptide (e.g., SEQ ID NO: 1).

The aryl sulfotransferase activity may for instance be determined in accordance to the following method: Aryl sulfotransferase activity may be determined by the reaction of radioactively sulfur labelled PAPS, [^{35}S]PAPS, with the substrate in presence of the polypeptide of interest. This is described previously, for example by Hattori *et al* (Biosci Biotechnol Biochem. 2008; 72(2):540-7). The reaction takes place in a buffer such as 250 µL 50 mM sodium phosphate pH 6.8 with 1 µM [^{35}S]PAPS (3.7kBq) with 100 µM accepting compound for a period of 30 min at 30°C. The reaction is stopped by addition of 100 µL of a 1:1 mixture of 0.1 M barium acetate and barium hydroxide. 50 µL of 0.1 M zinc sulfate is

added, followed by centrifugation at 1,200 × g for 5 min. 300 µL of the supernatant is then transferred to a new container and 50 µL of an equal volume of 0.1 M barium hydroxide and 0.1 M zinc sulfate is added. The mixture is then centrifuged at 13,000 × g for 5 min, and 300-µL aliquots of the supernatant are mixed with 2.5 mL of Cleasol I (Nacalai Tesque, Kyoto, Japan). The radioactivity is then measured by scintillation.

Alternatively, the activity of a sulfotransferase may be detected by direct measurement of the product by analytical methods such as high performance liquid chromatography (HPLC) or liquid chromatography in combination with mass spectrometry (LC-MS).

Sulfate supply may be a limiting factor in an enzymatic sulfation reaction, and hence in the production of sulfated phenolic compounds, such as zosteric acid. Here, the present inventors have demonstrated that the production of sulfated phenolic compounds can be significantly increased if the sulfate uptake by the recombinant host cell is increased.

Therefore, a recombinant host cell according to the invention may be one which has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

Sulfate uptake by a given cell may be determined by a S^{35}-sulfate based method as described, e.g., by Mansilla and Mendoza (Microbiology, 2000, **146**, 815-821). Generally, cells are first grown in a defined minimal medium, such as M9 minimal medium, supplemented with glutathione as sulphur source to exponential phase. Cells are collected, washed and then resuspended in minimal medium. The measurement of sulfate uptake is performed by incubating for 5 min at 30°C the cell suspension containing 10^8 cells ml^{-1}, 0,01 mM sodium sulfate and approximately 10^6 cpm $^{35}SO_4^{2-}$ ml^{-1} (1050 Ci $mmol^{-1}$). The incubation period is terminated by filtering the cell suspension through a 0,45 µm Millipore filter, followed by washing the filters with 5 ml minimal medium containing 2 mM magnesium sulfate and 2 mM sodium thiosulfate. Filters are transferred to polyethylene vials containing 2 ml Optiphase 'HiSafe 3' scintillation fluid (Wallac) and the radioactivity counted in an LKB Primo liquid scintillation counter. Uptake rates are expressed in nmol sulfate min^{-1} (g cellular protein)$^{-1}$.

More particularly, a recombinant host cell according to the present invention may be modified to have an increased protein expression of sulfate transporter compared to the identical host cell that does not carry said modification. By "increased protein expression" it

is meant that the amount of the sulfate transporter protein produced by the thus modified host cell is increased compared an identical host cell that does not carry said modification. More particularly, by "increased expression" it is meant that the amount of the sulfate transporter protein produced by the thus modified host cell is increased by at least 10%, such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700% at least 800%, at least about 900%, at least about 1000%, at least about 2000%, at least about 3000%, at least about 4000%, at least about 5000%, at least about 6000%, at least about 7000%, at least about 8000% at least about 9000% or at least about 10000%, compared an identical host cell that does not carry said modification. The amount of protein in a given cell can be determined by any suitable quantification technique known in the art, such as ELISA, Immunohistochemistry or Western Blotting.

An increase in protein expression may be achieved by any suitable means well-know to those skilled in the art. For example, an increase in protein expression may be achieved by increasing the number of copies of the gene or genes encoding the sulfate transporter in the host cell, such as by introducing into the host cell a exogenous nucleic acid, such as a vector, comprising the gene or genes encoding the sulfate transporter operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule. An increase in protein expression may also be achieved by integration of at least a second copy of the gene or genes encoding the sulfate transporter into the genome of the host cell. An increase in protein expression may also be achieved by increasing the strength of the promoter(s) operably linked to the gene or genes encoding the sulfate transporter. An increase in protein expression may also be achieved by modifying the ribosome binding site on the mRNA molecule encoding the sulfate transporter. By modifying the sequence of the ribosome binding site the translation initiation rate may be increased, thus increasing the translation efficiency.

According to certain embodiments, the increase in the number of copies of the gene or genes is achieved by introducing into the recombinant host cell one or more (such as two or three) exogenous nucleic acid molecules (such as one or more vectors) comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule.

According to certain embodiments, a recombinant host cell is provided which comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more (such as two, three or four) nucleotide sequences encoding a sulfate transporter. Suitably, the exogenous nucleic acid molecule further comprises a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the one or more nucleotide sequences encoding said sulfate transporter.

According to certain embodiments, the exogenous nucleic acid molecule is stably integrated into the genome of the recombinant host cell.

The sulfate transporter may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the sulfate transporter is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the sulfate transporter is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to certain embodiments, the sulfate transporter is a bacterial sulfate transporter. With "bacterial sulfate transporter" it is meant that the sulfate transporter is naturally derived from a bacterium, such as *Escherichia coli*.

The sulfate transporter employed in accordance of the invention may be any suitable sulfate transporter which is functional in the respective host cell.

According to certain embodiments, the sulfate transporter is a selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

Accoriding to certain embodiments, the sulfate transporter is a bacterial sulfate transporter selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

According to particular embodiments, the sulfate transporter is a CysZ protein.

Members of the CysZ family (TCDB 2.A.121) are high affinity, high specificity proton-dependent sulfate transporters which mediates sulfate uptake. Non-limiting examples of CysZ proteins are those found in bacteria, such as *E. coli* (NCBI: NP_416908.1) *S. typhimurium* (NCBI: NP_456966.1), *K. pneumoniae* (NCBI: CDO15722.1), *P. fluorescens* (NCBI: AEV64873.1), *S. sonnei* (NCBI: AAZ89133.1), *V. anguillarum* (NCBI: AEH33702.1), *B. japonicum* (NCBI: KOY11972.1) and *C. glutamicum* (NCBI: CAF20834.1) to only name a few.

Accordingly, a sulfate transporter for use according to the invention may for instance be the CysZ protein from *Escherichia coli* (SEQ ID NO: 14). Further information regarding CysZ of Escherichia coli is available at EcoCyc (www.biocyc.org) under Accession number EG10003. See also NCBI Reference Sequence Database under NCBI Reference Sequence: NP_416908.1.

According to certain embodiments, the sulfate transporter is a polypeptide selected from the group consisting of:

2a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14;

2b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14; or

2c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 2a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 2b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has

at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14. According to other particular embodiments, a polypeptide according to 2b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 14.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 2c). According to particular embodiments, a polypeptide according to 2c) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 2c) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 2c) comprises an amino acid sequence set forth in SEQ ID NO: 14, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 14). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be

amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 2b) or 2c) has sulfate transporter activity. More preferably, a polypeptide according to 2b) or 2c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 14. With "similar" sulfate transporter activity it is meant that the polypeptide according to 2b) or 2c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 14).

Another suitable sulfate transporter for use according to the invention may for instance be the CysZ protein from *Corynebacterium glutamicum* (SEQ ID NO: 15). Further information regarding CysZ of *C. glutamicum* is available at NCBI under accession number CAF20834.1.

According to certain embodiments, the sulfate transporter is a polypeptide selected from the group consisting of:

3a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15;

3b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15; or

3c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 3a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 3b). Accordingly, a sulfate transporter employed according to the present invention is a

polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to other particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to other particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15. According to other particular embodiments, a polypeptide according to 3b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 15.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 3c). According to particular embodiments, a polypeptide according to 3c) comprises an amino acid sequence set forth in SEQ ID NO: 15, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 3c) comprises an amino acid sequence set forth in SEQ ID NO: 15, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 3c) comprises an amino acid sequence set forth in SEQ ID NO: 15, wherein about 1 to about 25, such as about 1 to about

20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 15). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 3b) or 3c) has sulfate transporter activity. More preferably, a polypeptide according to 3b) or 3c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 15. With "similar" sulfate transporter activity it is meant that the polypeptide according to 3b) or 3c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 15).

According to certain embodiments, the sulfate transporter is a sulfate-transporting ATPase, such as a member of the SulT (cysPTWA) family.

According to particular embodiments, the sulfate transporter is bacterial sulfate transporter of the SulT (cysPTWA) family.

Sulfate transporters of the SulT (cysPTWA) family from proteobacteria (TCDB 3.A.1.6.1), such as *E. coli*, are generally constituted by: (i) one of two periplasmic proteins, SbP, the sulfate binding protein, or CysP, the thiosulfate-binding protein; (ii) membrane proteins CysT (synonym: CysU) and CysW; and (iii) the ATP-binding protein CysA. The SulT subunits are encoded by the cysPTWA operon and by the sbp gene, located either in another chromosomal region or instead of cysP in the same operon. Non-limiting examples of sulfate transporters of the SulT (cysPTWA) family are those found in *Escherichia coli, Salmonella typhimurium* and *Rhodobacter capsulatus*.

CysT, CysW, CysA, CysP and Sbp of *Escherichia coli* are set forth in SEQ ID NO: 16 to 20, respectively. Further information regarding CysT, CysW, CysA, CysP and Sbp of *Escherichia coli* is available at EcoCyc (www.biocyc.org) under Accession numbers EG10197, EG10198, EG10183, EG10195 and EG10929, respectively.

CysT, CysW, CysA, CysP and Sbp of *Salmonella typhimurium* are set forth in SEQ ID NO: 21 to 25, respectively.

According to certain embodiments, the sulfate transporter comprises a first membrane subunit (CysT), a second membrane subunit (CysW), an ATP binding subunit (CysA) and a periplasmic binding protein (CysP or Sbp).

According to certain embodiments, the sulfate transporter is encoded by an operon comprising a nucleotide sequence encoding a first membrane subunit (CysT), a nucleotide sequence encoding a second membrane subunit (CysW), a nucleotide sequence encoding an ATP binding subunit (CysA) and a nucleotide sequence encoding a periplasmic binding protein (CysP or Sbp).

According to certain embodiments, the first membrane subunit is a polypeptide selected from the group consisting of:

4a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16 or 21;

4b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16 or 21; or

4c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16 or 21, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the first membrane subunit is a polypeptide according to 4a). According to particular embodiments, the first membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, the first membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 21.

According to other certain embodiments, the first membrane subunit is a polypeptide according to 4b). Accordingly, a first membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 16.

Alternatively, a first membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the

amino acid sequence set forth in SEQ ID NO: 21. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21. According to other particular embodiments, a polypeptide according to 4b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 21.

According to other certain embodiments, the first membrane subunit is a polypeptide according to 4c). According to particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 16, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 16, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 16, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 21, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 4c) comprises an amino acid sequence set forth in SEQ ID NO: 21, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 4c) comprises an amino acid

sequence set forth in SEQ ID NO: 21, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 16 or 21). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 4b) or 4c) assembles with the proteins CysW, CysA and CysP/Sbp to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the second membrane subunit is a polypeptide selected from the group consisting of:

5a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17 or 22;

5b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17 or 22; or

5c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17 or 22, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the second membrane subunit is a polypeptide according to 5a). According to particular embodiments, the second membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, the second membrane subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 22.

According to other certain embodiments, the second membrane subunit is a polypeptide according to 5b). Accordingly, a second membrane subunit may be a polypeptide

comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 17.

Alternatively, a second membrane subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to other particular

embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22. According to other particular embodiments, a polypeptide according to 5b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 22.

According to other certain embodiments, the second membrane subunit is a polypeptide according to 5c). According to particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 17, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 17, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 17, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 22, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 22, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 5c) comprises an amino acid sequence set forth in SEQ ID NO: 22, wherein about 1 to about 25, such as about 1 to about

20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 17 or 22). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 5b) or 5c) assembles with the proteins CysT, CysA and CysP/Sbp to form a sulfate-transporting ATPase which tranports sulfate into the host cell.

According to certain embodiments, the ATP binding subunit is a polypeptide selected from the group consisting of:

6a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18 or 23;

6b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18 or 23; or

6c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18 or 23, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the ATP binding subunit is a polypeptide according to 6a). According to particular embodiments, the ATP binding subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, the ATP binding subunit is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 23.

According to other certain embodiments, the ATP binding subunit is a polypeptide according to 6b). Accordingly, a ATP binding subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least

about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 18.

Alternatively, a ATP binding subunit may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at

least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23. According to other particular embodiments, a polypeptide according to 6b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 23.

According to other certain embodiments, the ATP binding subunit is a polypeptide according to 6c). According to particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 18, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 18, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 18, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 23, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 23, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 6c) comprises an amino acid sequence set forth in SEQ ID NO: 23, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 18 or 23). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 6b) or 6c) assembles with the proteins CysT, CysW and CysP/Sbp to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the periplasmic binding protein is CysP.

According to certain embodiments, the periplasmic binding protein is a polypeptide selected from the group consisting of:

7a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19 or 24;

7b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19 or 24; or

7c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19 or 24, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the periplasmic binding protein is a polypeptide according to 7a). According to particular embodiments the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 24.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 7b). Accordingly, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least

95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 19.

Accordingly, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at

least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24. According to other particular embodiments, a polypeptide according to 7b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 24.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 7c). According to particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 19, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 19, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 19, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 24, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 24, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 7c) comprises an amino acid sequence set forth in SEQ ID NO: 24, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 19 or 24). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 7b) or 7c) assembles with the proteins CysT, CysW and CysA to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the periplasmic binding protein is Sbp.

According to certain embodiments, the periplasmic protein is a polypeptide selected from the group consisting of:

8a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20 or 25;

8b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20 or 25; or

8c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20 or 25, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the periplasmic binding protein is a polypeptide according to 8a). According to particular embodiments the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, the periplasmic binding protein is a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 25.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 8b). Accordingly, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the

amino acid sequence set forth in SEQ ID NO: 20. According to particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 20.

Alternatively, a periplasmic binding protein may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25. According to particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino

acid sequence set forth in SEQ ID NO: 25. According to other particular embodiments, a polypeptide according to 8b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 25.

According to other certain embodiments, the periplasmic binding protein is a polypeptide according to 8c). According to particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 20, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 20, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 20, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

According to other particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 25, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 25, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 8c) comprises an amino acid sequence set forth in SEQ ID NO: 25, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 20 or 25). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, the polypeptide according to 8b) or 8c) assembles with the proteins CysT, CysW and CysA to form a sulfate-transporting ATPase which transports sulfate into the host cell.

According to certain embodiments, the sulfate transporter is a member of the SulP family.

The sulfate transporter SulP family (TCDB 2.A.53) is a large and ubiquitous family with members derived from archaea, bacteria, fungi, plants and animals. Many organisms including *Bacillus subtilis*, *Synechocystis sp*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Caenorhabditis elegans* possess multiple SulP family paralogues. Many of these proteins are functionally characterized, and most are inorganic anion uptake transporters or anion:anion exchange transporters. A non-limiting example of a sulfate transporter of the SulP family is that found in *Mycobacterium tuberculosis* (SEQ ID NO: 26; NCBI: NP_216255.1). Zolotarev et al. (Comp Biochem Physiol A Mol Integr Physiol. 2008 Mar; 149(3):255-66) have demonstrate that the overexpression of SulP protein Rv1739c from *M. tuberculosis* in *E. coli* increases sulfate uptake. Another non-limiting example is a SulP protein found in multiple species (NCBI: WP_012536065.1).

According to certain embodiments, the sulfate transporter is a polypeptide selected from the group consisting of:

9a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26;

9b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26; or

9c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 9a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 9b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to other particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to other particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26. According to other particular embodiments, a polypeptide according to 9b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 26.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 9c). According to particular embodiments, a polypeptide according to 9c) comprises an amino acid sequence set forth in SEQ ID NO: 26, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 9c) comprises an amino acid sequence set forth in SEQ ID NO: 26, wherein about 1 to about 30, such as about 1 to about 25, about

1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 9c) comprises an amino acid sequence set forth in SEQ ID NO: 26, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 26). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 9b) or 9c) has sulfate transporter activity. More preferably, a polypeptide according to 9b) or 9c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 26. With "similar" sulfate transporter activity it is meant that the polypeptide according to 9b) or 9c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 26).

According to certain embodiments, the sulfate transporter is a CysP transporter belonging to the phosphate inorganic transporter (PiT) family.

Genes encoding PiT family transporters are widespread throughout the three life domains. A non-limiting example of a CysP transporter of the PiT family is that found in *Bacillus subtilis* (SEQ ID NO: 27; GenBank: CAB13432.1). Another non-limiting example of a CysP transporter of the PiT family is that found in *Halobacterium salinarum* (GenBank: CAP13497.1).

According to certain embodiments, the sulfate transporter is selected from the group consisting of:

10a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27;

10b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27; or

10c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 10a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 10b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to other particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to other particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27. According to other particular embodiments, a polypeptide according to 10b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at

least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 27.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 10c). According to particular embodiments, a polypeptide according to 10c) comprises an amino acid sequence set forth in SEQ ID NO: 27, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 10c) comprises an amino acid sequence set forth in SEQ ID NO: 27, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 10c) comprises an amino acid sequence set forth in SEQ ID NO: 27, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 27). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 10b) or 10c) has sulfate transporter activity. More preferably, a polypeptide according to 10b) or 10c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 27. With “similar” sulfate transporter activity it is meant that the polypeptide according to 10b) or 10c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 27).

According to certain embodiments, the sulfate transporter is an oxyanion permease (PerO).

Oxyanion permeases act as a general oxyanion importer of molybdate, sulfate, tungstate, and vanadate. A non-limiting example of an oxyanion permease is that found in *Rhodobacter capsulatus* (SEQ DI NO: 95).

According to certain embodiments, the sulfate transporter is selected from the group consisting of:

12a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 95;

12b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 95; or

12c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 95, wherein 1 to 50, such as 1 to 40, 1 to 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the sulfate transporter is a polypeptide according to 12a).

According to other certain embodiments, the sulfate transporter is a polypeptide according to 12b). Accordingly, a sulfate transporter employed according to the present invention is a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 95. According to particular embodiments, a polypeptide according to 12b) comprises an amino acid sequence which has at least about 80%, such as at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 95. According to other particular embodiments, a polypeptide according to 12b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO:

95. According to other particular embodiments, a polypeptide according to 12b) comprises an amino acid sequence which has at least about 90%, such as at least about 93%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 95. According to other particular embodiments, a polypeptide according to 12b) comprises an amino acid sequence which has at least about 95%, such as at least about 96%, at least about 97%, at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 95.

According to other certain embodiments, the sulfate transporter is a polypeptide according to 12c). According to particular embodiments, a polypeptide according to 12c) comprises an amino acid sequence set forth in SEQ ID NO: 95, wherein about 1 to about 50, such as about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to more particular embodiments, a polypeptide according to 12c) comprises an amino acid sequence set forth in SEQ ID NO: 95, wherein about 1 to about 30, such as about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted. According to other more particular embodiments, a polypeptide according to 12c) comprises an amino acid sequence set forth in SEQ ID NO: 95, wherein about 1 to about 25, such as about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (i.e., SEQ ID NO: 95). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 12b) or 12c) has sulfate transporter activity. More preferably, a polypeptide according to 12b) or 12c) has a sulfate transporter activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 95. With “similar” sulfate transporter activity it is meant that the polypeptide according to 12b) or 12c) has at least about 10%, such as at least about 20%, at least about 30%, at least

about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the sulfate transporter activity of the reference polypeptide (i.e., SEQ ID NO: 95).

Besides external sulfate supply, the sulfation reaction may further depend on the supply of sulfate from 3′ -phosphoadenosine 5′ -phosphosulfate (PAPS) or transferred from another sulfated compound. The inventors have shown that the sulfation reaction can further be enhanced by improving the supply of PAPS (3' -phosphoadenosine 5' -phosphosulfate) and, in addition, by the removal of the product 3′ -phosphoadenosine 5′ -phosphate (PAP). The improved supply is obtained by deregulation, mutation or overexpression of enzymes that increase PAPS concentration or similarly reduce PAP concentration. This is exemplified in Example 2, where an increased production of zosteric acid in *Escherichia coli* is obtained by increasing the expression of the genes cysD, cysN, and cysC which are responsible for production of PAPS. Without being bound to a specific theory, it is believed that an adenylyl moiety (AMP) of ATP is transferred to sulfate to form activated sulfate, or APS (adenosine 5'-phosphosulfate). This extremely unfavorable reaction is kinetically and energetically linked to the hydrolysis of GTP by the enzyme ATP sulfurylase, which is composed of two types of subunits: an adenylyl transferase (cysD) and a GTPase (cysN). APS is then phosphorylated at the 3'-hydroxyl to form PAPS (3'-phosphoadenosine 5'-phosphosulfate) in a reaction catalysed by APS kinase, which is encoded by cysC. Furthermore, the inventors have enhanced the production of zosteric acid even more by increasing the expression of the gene cysQ encoding a PAP phosphatase which is responsible for the removal of PAP.

Therefore, in order to further improve the production of a sulfated phenolic compound, such as zosteric acid, a recombinant host cell according to the present invention may be further modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification; may be further modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification; and/or may be further modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification. By "increased protein expression" it is meant that the amount of

the respective protein produced by the thus modified host cell is increased compared an identical host cell that does not carry said modification. More particularly, by "increase expression" it is meant that the amount of respective protein produced by the thus modified host cell is increased by at least 10% , such as at least 20%, at least 30%, at least 40%, at least 50% at least 60%, at least 70%, at least 80%, at least 90%, at least 100%, at least 150%, at least 200%, at least 300%, at least 400%, at least 500%, at least 600%, at least 700% at least 800%, at least about 900%, at least about 1000%, at least about 2000%, at least about 3000%, at least about 4000%, at least about 5000%, at least about 6000%, at least about 7000%, at least about 8000% at least about 9000% or at least about 10000%, compared an identical host cell that does not carry said modification. The amount of protein in a given cell can be determined by any suitable quantification technique known in the art, such as ELISA, Immunohistochemistry or Western Blotting.

According to certain embodiments, a recombinant host cell according to the invention has further been modified to have an increased protein expression an ATP sulfurylase compared to an identical host cell that does not carry said modification.

According to certain embodiments, a recombinant host cell according to the invention has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification.

According to certain embodiments, a recombinant host cell according to the invention has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

An increase in protein expression may be achieved by any suitable means well-know to those skilled in the art. For example, an increase in protein expression may be achieved by increasing the number of copies of the gene or genes encoding the respective protein (e.g., ATP sulfurylase, APS kinase and/or PAP phosphatase) in the host cell, such as by using (e.g., introducing into the host cell) a vectors comprising the gene or genes operably linked to a promoter that is functional in the host cell to cause the production of an mRNA molecule. An increase in protein expression may also be achieved by integration of at least a second copy of the gene or genes encoding the respective protein into the genome of the host cell. An increase in protein expression may also be achieved by increasing the strength of the promoter(s) operably linked to the gene or genes. An increase in protein expression may

also be achieved by modifying the ribosome binding site on the mRNA molecule encoding the respective protein (e.g., ATP sulfurylase, APS kinase and/or PAP phosphatase). By modifying the sequence of the ribosome binding site the translation initiation rate may be increased, thus increasing the translation efficiency.

ATP sulfurylase encoding genes for use according to the invention may for instance be the cysD and cysN genes from *Escherichia coli* (encoding SEQ ID NO: 28 and 29, respectively). Alternative ATP sulfurylase encoding genes include the *Arabidopsis thaliana* ATP sulfurylase ASAL gene (GenBank Accession No. U40715, Logan et al. (1996) J Biol Chem 271: 12227); the *Allium cepa* ATP-sulfurylase gene (GenBank Accession No AF21154); the *Lotus japonicus* ATP sulfurylase gene (GenBank Accession No. AW164083); the *Arabidopsis thaliana* met3-1 ATP sulfurylase gene (GenBank Accession No. X79210).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising one or more nucleotide sequences encoding a ATP sulfurylase.

The ATP sulfurylase may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the ATP sulfurylase is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the ATP sulfurylase is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to certain embodiments, the ATP sulfurylase is a protein constituted by two polypeptides, which are exemplified by the amino acid sequence set forth in SEQ ID NO: 28 and 29, respectively.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28, and a nucleotide sequence encoding iii) a polypeptide comprising an amino acid sequence set forth in SEQ ID

NO: 29 or iv) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 28 and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 29.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28 and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 28 and a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 29. Preferably, the polypeptides assemble to form a protein having ATP sulfurylase activity.

An alternative ATP sulfurylase encoding gene for use according to the invention may for instance be the MET3 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 30).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 30 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 30. Preferably, the polypeptide according to ii) has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 30.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 30. Preferably, the polypeptide has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 30. Preferably, the polypeptide has ATP sulfurylase activity.

An alternative ATP sulfurylase encoding gene for use according to the invention may for instance be the ATP sulfurylase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 31).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 31 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 31. Preferably, the polypeptide according to ii) has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 31.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 31. Preferably, the polypeptide has ATP sulfurylase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 31. Preferably, the polypeptide has ATP sulfurylase activity.

Techniques for determining ATP sulfurylase activity are well known to the skilled person. Exemplary methods have been described, e.g. by Reuveny and Filner (Anal Biochem, 1976, 75(2), 410-428) or Hommes and Moss (Anal Biochem, 1986, 154(1), 100-103).

An APS kinase encoding gene for use according to the invention may for instance be the cysC gene from *Escherichia coli* (encoding SEQ ID NO: 32).

In certain instances a single polypeptide has been shown to possess both an ATP sulfurylase and a 5'-adenylylsulfate kinase activity. For example, an ATP sulfurylase/APS kinase encoding gene has been isolated from mouse (GenBank Accession No. U34883, Li et al. (1995) J Biol Chem)70: 1945), and human (GenBank Accession No. AF033026, Yanagisawa (1998) Biosci Biotechnol Biochem 62: 1037) sources. Other examples of such bifunctional enzyme include 3'-phosphoadenosine 5'-phosphosulfate synthase enzymes (PAPSS) from rat (*Rattus norvegicus*) (SEQ ID NO: 33 or 34).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an APS kinase.

The APS kinase may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the APS kinase is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the APS kinase is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 32.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least

about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 32. Preferably, said polypeptide has APS kinase activity.

An alternative APS kinase encoding gene for use according to the invention may for instance be the MET14 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 35).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 35 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 35. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 35.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 35. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 35. Preferably, said polypeptide has APS kinase activity.

An alternative APS kinase encoding gene for use according to the invention may for instance be the APS kinase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 36).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 36 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 36. Preferably, said polypeptide according to ii) has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 36.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 36. Preferably, said polypeptide has APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at

least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 36. Preferably, said polypeptide has APS kinase activity.

Alternatively, a polypeptide having both an ATP sulfurylase and a APS kinase activity can be used, such as a 3'-phosphoadenosine 5'-phosphosulfate synthase (PAPSS).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an 3'-phosphoadenosine 5'-phosphosulfate synthase.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 33 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 33. Preferably, said polypeptide according to ii) has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 33.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 33. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at

least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 33. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 34 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 34. Preferably, said polypeptide according to ii) has both an ATP sulfurylase and a APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 34.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 34. Preferably, said polypeptide has both an ATP sulfurylase and APS kinase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 34. Preferably, said polypeptide has both an ATP sulfurylase and a APS kinase activity.

Techniques for determining APS kinase activity are well known to the skilled person. An exemplary method has been described, e.g. by Burnell and Whatley (Anal Biochem, 1975, 68(1), 281-288).

A PAP phosphatase encoding gene for use according to the invention may for instance be the cysQ gene from *Escherichia coli* (encoding SEQ ID NO: 37).

According to certain embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding an PAP phosphatase.

The PAP phosphatase may be derived from the same species as the recombinant host cell in which it is expressed or may be derived from a species different to the one in which it is expressed (i.e. it is heterologous). According to certain embodiments, the PAP phosphatase is derived from the same species as the recombinant host cell in which it is expressed. According to certain other embodiments, the PAP phosphatase is derived from a species different to the one in which it is expressed (i.e. it is heterologous).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 37.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least

about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 37. Preferably, said polypeptide has PAP phosphatase activity.

An alternative PAP phosphatase encoding gene for use according to the invention may for instance be the MET22 gene from *Saccharomyces cerevisiae* (encoding SEQ ID NO: 38).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 38 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 38. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 38.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 38. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 38. Preferably, said polypeptide has PAP phosphatase activity.

An alternative PAP phosphatase encoding gene for use according to the invention may for instance be the PAP phosphatase encoding gene from *Bacillus subtilis* (encoding SEQ ID NO: 39).

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding i) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 39 or ii) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 39. Preferably, said polypeptide according to ii) has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as a vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 39.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 39. Preferably, said polypeptide has PAP phosphatase activity.

According to particular embodiments, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule (such as vector) comprising a nucleotide sequence encoding a polypeptide comprising an amino acid sequence which has at least

about 85%, such as at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 39. Preferably, said polypeptide has PAP phosphatase activity.

Techniques for determining PAP phosphatase activity are well known to the skilled person. An exemplary method has been described, e.g. by Fukuda et al. (Appl Environ Microbiol, 2007, 73(17), 5447-5452).

According to certain embodiments, the nucleotide sequences coding for an ATP sulfurylase, an APS kinase and an PAP phosphatase, respectively, are part of an operon. Accordingly, a recombinant host cell according to the invention may comprise an exogenous nucleic acid molecule (such as vector) which comprises an operon comprising a nucleotide sequence or nucleotide sequences encoding an ATP sulfurylase, a nucleotide sequence encoding an APS kinase, and optionally a nucleotide sequence encoding an PAP phosphatase.

Contemplated by the present invention is the production of a sulfated phenolic compound from a precursor thereof, and in particular from a precursor of the general formula (p-I) as described in more detail below. In this case, it may be suitable to employ (e.g. to express by a host cell of the invention) a polypeptide which has tyrosine ammonia lyase activity.

Tyrosine ammonia lyases suitable for use according to the present invention have been described in, e.g., WO 2016/008886 A1.

Therefore, a recombinant host cell provided and utilized in accordance with the present invention may comprise a hetereologus polypeptide having tyrosine ammonia lyase activity. According to certain embodiments, a recombinant host cell according to the invention comprises a heterologous polypeptide selected from the group consisting of:

11a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40);

11b) a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99%, sequence

identity to the amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40); or

11c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40), wherein 1 or more, such as about 1 to about 50, about 1 to about 40, about 1 to about 35, 1 to 30, 1 to 25, 1 to 20, 1 to 15, 1 to 10, 1 to 5, 1 to 3, amino acid residues are substituted, deleted and/or inserted.

According to certain embodiments, the polypeptide having tyrosine ammonia lyase activity is a polypeptide according to 11a). Accordingly, a polypeptide having tyrosine ammonia lyase activity may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40).

According to certain embodiments, the polypeptide having tyrosine ammonia lyase activity is a polypeptide according to 11b). Accordingly, a polypeptide having tyrosine ammonia lyase activity may be a polypeptide comprising an amino acid sequence which has at least about 70%, such as at least about 75%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40). According to other particular embodiments, a polypeptide according to 11b) comprises an amino acid sequence which has at least about 85%, such as at least about 90%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40). According to other particular embodiments, a polypeptide according to 11b) comprises an amino acid sequence which has at least about 95%, such as at least about 98%, or at least about 99%, sequence identity to the amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40).

According to certain embodiments, the polypeptide having tyrosine ammonia lyase activity is a polypeptide according to 11c). Accordingly, a polypeptide having tyrosine ammonia lyase activity may be a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40), wherein 1 or more, such as 2 or more, 3 or more, 4 or more, 5 or more, 6 or more, 7 or more, 8 or more, 9 or more, 10 or more, 11 or more, 12 or more, 13 or more, 14 or more, 15 or more, 16 or more, 17 or more, 18 or more, 19 or more, 20 or more, 25 or more, 30 or more, 35 or more, 40 or more, 45 or more, 50 or more, 60 or more, 70 or more, 80 or more, 90 or more, 100 or

more, 110 or more, 120 or more, 130 or more, 140 or more, or 150 or more, amino acid residues are substituted, deleted, and/or inserted.

According to particular embodiments, a polypeptide according to 11c) comprises an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40), wherein about 1 to about 50, about 1 to about 40, about 1 to about 35, about 1 to about 30, about 1 to about 25, about 1 to about 20, about 1 to about 15, about 1 to about 10, about 1 to about 5, or about 1 to about 3, amino acid residues are substituted, deleted and/or inserted.

It is understood that the foregoing values generally define the total number of alterations to the reference polypeptide (e.g., SEQ ID NO: 40). The alterations may solely be amino acid substitutions, be it conserved or non-conserved substitutions, or both. They may solely be amino acid deletions. They may solely be amino acid insertions. The alterations may be a mix of these specific alterations, such as amino acid substitutions and amino acid insertions.

Preferably, a polypeptide according to 11b) or 11c) has tyrosine ammonia lyase activity. More preferably, a polypeptide according to 11b) or 11c) has a tyrosine ammonia lyase activity similar to that of the polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 (e.g., SEQ ID NO: 40). According to certain embodiment, a polypeptide according to 11b) or 11c) has tyrosine ammonia lyase activity similar to that of the polypeptide comprising the amino acid sequence set forth in SEQ ID NO: 40. With “similar” tyrosine ammonia lyase activity it is meant that the polypeptide according to 11b) or 11c) has at least about 10%, such as at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60, at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 100%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 800%, at least about 1000% or at least about 2000%, of the tyrosine ammonia lyase activity of the reference polypeptide (e.g., SEQ ID NO: 40). Tyrosine ammonia lyase activity may be determined according to the method described in WO 2016/008886 A1 (e.g., page 9, line 29 to page 10, line 2).

Alternatively, the heterologous polypeptide having aryl sulfotransferase activity may be comprised by a first recombinant host cell, and the heterologous polypeptide having tyrosine ammonia lyase activity may be comprised by a second recombinant host cell.

Further contemplated by the present invention is to employ a further (e.g., third) heterologous polypeptide which has phenylalanine ammonia lyase activity, such as a phenylalanine ammonia lyase (EC 4.3.1.24).

According to certain embodiments, a recombinant host cell comprises (e.g., expresses) a heterologous polypeptide having aryl sulfotransferase activity and a heterologous polypeptide having phenylalanine ammonia lyase activity.

Alternatively, the heterologous polypeptide having aryl sulfotransferase activity may be comprised by a first recombinant host cell, and the heterologous polypeptide having phenylalanine ammonia lyase activity may be comprised by a furhter recombinant host cell. Such further recombinant host cell may be a recombinant host cell also comprising a heterologous polypeptide having tyrosine ammonia lyase activity.

Recombinant host cells in accordance with the invention can be produced from any suitable host organism, including single-celled or multicellular microorganisms such as bacteria, yeast, fungi, algae and plant, and higher eukaryotic organisms including nematodes, insects, reptiles, birds, amphibians and mammals.

According to certain embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, algae and plant.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast, fungi, and algae.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria, yeast and fungi.

According to certain other embodiments, a recombinant host cells in accordance with the invention is selected from the group consisting of bacteria and yeast.

According to certain embodiments, a recombinant host cells in accordance with the invention is not a plant cell.

Bacterial host cells are selected from Gram-positive and Gram-negative bacteria. Non-limiting examples for Gram-negative bacterial host cells include species from the genera *Escherichia*, *Erwinia*, *Klebsiella* and *Citrobacter*. Non-limiting examples of Gram-positive

bacterial host cells include species from the genera *Bacillus*, *Lactococcus*, *Lactobacillus*, *Clostridium*, *Corynebacterium*, *Streptomyces*, *Streptococcus*, and *Cellulomonas*.

According to certain embodiment, the recombinant host cell is a bacterium of the family selected from the group consisting of *Enterobacteriaceae, Bacillaceae, Lactobacillaceae* and *Corynebacteriaceae.* According to certain embodiments, the recombinant host cell is a bacterium of the family *Enterobacteriaceae.*

According to certain embodiments, the recombinant host cell is a bacterium, which may be a bacterium of the genus *Bacillus, Lactococcus, Lactobacillus, Clostridium, Corynebacterium, Geobacillus, Thermoanaerobacterium, Streptococcus, Pseudomonas, Streptomyces, Escherichia, Shigella, Acinetobacter, Citrobacter, Salmonella, Klebsiella, Enterobacter, Erwinia, Kluyvera, Serratia, Cedecea, Morganella, Hafnia, Edwardsiella, Providencia, Proteus*, or *Yersinia*.

According to particular embodiments, the recombinant host cell is a bacterium of the genus Bacillus. Non-limiting examples of a bacteria of the genus *Bacillus* are *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, and *Bacillus mojavensis*. According to more particular embodiments, the recombinant host cell is *Bacillus subtilis*. According to other more particular embodiments, the recombinant host cell is *Bacillus licheniformis*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Lactococcus*. A non-limiting example of a bacterium of the genus *Lactococcus* is *Lactococcus lactis*. According to more particular embodiments, the recombinant host cell is *Lactococcus lactis*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Corynebacterium*. A non-limiting example of a bacterium of the genus *Corynebacterium* is *Corynebacterium glutamicum*. According to more particular embodiments, the recombinant host cell is *Corynebacterium glutamicum*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Streptomyces*. A non-limiting examples of a bacterium of the genus *Streptomyces* are *Streptomyces lividans*, *Streptomyces coelicolor*, or *Streptomyces griseus*. According to more particular embodiments, the recombinant host cell is *Streptomyces lividans*. According to other more particular embodiments, the recombinant host cell is *Streptomyces coelicolor*.

According to other more particular embodiments,, the recombinant host cell is *Streptomyces griseus*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Pseudomonas. A non-limiting example of a bacterium of the genus *Pseudomonas* is *Pseudomonas putida*. According to more particular embodiments, the recombinant host cell is *Pseudomonas putida*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Geobacillus*. A non-limiting examples of a bacterium of the genus *Geobacillus* are *Geobacillus thermoglucosidasius* and *Geobacillus stearothermophilus*. According to more particular embodiments, the recombinant host cell is *Geobacillus thermoglucosidasius*. According to other more particular embodiments, the recombinant host cell is *Geobacillus stearothermophilus*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus Thermoanaerobacterium. A non-limiting example of a bacterium of the genus Thermoanaerobacterium is *Thermoanaerobacterium thermosaccharolyticum*. According to more particular embodiments, the recombinant host cell is *Thermoanaerobacterium thermosaccharolyticum*.

According to other particular embodiments, the recombinant host cell is a bacterium of the genus *Escherichia*. A non-limiting example of a bacterium of the genus Escherichia is *Escherichia coli*. According to more particular embodiments, the recombinant host cell is *Escherichia coli*.

Yeast host cells may be derived from e.g., *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

According to certain embodiments, the recombinant host cell is a yeast, which may be a yeast is of the genus *Saccharomyces, Pichia, Schizosacharomyces, Zygosaccharomyces, Hansenula, Pachyosolen, Kluyveromyces, Debaryomyces, Yarrowia, Candida, Cryptococcus, Komagataella, Lipomyces, Rhodospiridium, Rhodotorula*, or *Trichosporon*.

According to particular embodiments, the recombinant host cell is a yeast of the genus *Saccharomyces*. A non-limiting example of a yeast of the genus *Saccharomyces* is *Saccharomyces cerevisiae*. According to more particular embodiments, the recombinant host cell is *Saccharomyces cerevisiae*.

According to particular embodiments, the recombinant host cell is a yeast of the genus *Pichia*. Non-limiting example of a yeast of the genus *Pichia* are Pichia pastoris and *Pichia kudriavzevii*. According to more particular embodiments, the recombinant host cell is *Pichia pastoris*. According to other more particular embodiments, the recombinant host cell is *Pichia kudriavzevii*.

Fungi host cells may be derived from, e.g., *Aspergillus*.

According to certain embodiments, the recombinant host cell is a fungus, such as a fungi of the genus *Aspergillus*. Non-limiting examples of a fungus of the genus *Aspergillus* are *Aspergillus Oryzae*, *Aspergillus niger* or *Aspergillus awamsii*. According to more particular embodiments, the recombinant host cell is *Aspergillus Oryzae*. According to other more particular embodiments, the recombinant host cell is *Aspergillus niger*. According to other more particular embodiments, the recombinant host cell is *Aspergillus awamsii*.

Algae host cells may be derived from, e.g., *Chlamydomonas*, *Haematococcus*, *Phaedactylum*, *Volvox* or *Dunaliella*.

According to certain embodiments, the recombinant host cell is an alga, which may be an algae of the genus *Chlamydomonas*, *Haematococcus*, *Phaedactylum*, *Volvox* or *Dunaliella*.

According to particular embodiments, the recombinant host cell is an alga cell of the genus *Chlamydomonas*. A non-limiting example of an alga of the genus *Chlamydomonas* is *Chlamydomonas reinhardtii*.

According to particular embodiments, the recombinant host cell is an alga cell of the genus *Haematococcus*. A non-limiting example of an alga of the genus *Haematococcus* is *Haematococcus pluvialis*.

According to other particular embodiments, the recombinant host cell is an alga cell of the genus *Phaedactylum*. A non-limiting example of an alga of the genus *Phaedactylum* is *Phaedactylum tricornatum*.

A plant host cell may be derived from, e.g., soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

According to certain embodiments, the recombinant host cell is a plant cell, such as a plant cell selected from the group consisting of soybean, rapeseed, sunflower, cotton, corn, tobacco, alfalfa, wheat, barley, oats, sorghum, lettuce, rice, broccoli, cauliflower, cabbage, parsnips, melons, carrots, celery, parsley, tomatoes, potatoes, strawberries, peanuts, grapes, grass seed crops, sugar beets, sugar cane, beans, peas, rye, flax, hardwood trees, softwood trees, and forage grasses.

According to certain embodiments, a recombinant host cell according to the invention does not express an endogenous PAPS-dependent aryl sulfotransferase.

Generally, a recombinant host cell according to the invention has been genetically modified to express one or more polypeptides as detailed herein, which means that one or more exogenous nucleic acid molecules, such as DNA molecules, which comprise(s) a nucleotide sequence or nucleotide sequences encoding said polypeptide or polypeptides has been introduced in the host cell. Techniques for introducing exogenous nucleic acid molecule, such as a DNA molecule, into the various host cells are well-known to those of skill in the art, and include transformation (e.g., heat shock or natural transformation), transfection, conjugation, electroporation, microinjection and microparticle bombardment.

Accordingly, a recombinant host cell according to the invention comprises an exogenous nucleic acid molecule comprising a nucleotide sequence encoding a polypeptide as detailed herein.

In order to facilitate expression of a polypeptide in the host cell, the exogenous nucleic acid molecule may comprise suitable regulatory elements such as a promoter that is functional in the host cell to cause the production of an mRNA molecule and that is operably linked to the nucleotide sequence encoding said polypeptide.

Promoters useful in accordance with the invention are any known promoters that are functional in a given host cell to cause the production of an mRNA molecule. Many such

promoters are known to the skilled person. Such promoters include promoters normally associated with other genes, and/or promoters isolated from any bacteria, yeast, fungi, alga or plant cell. The use of promoters for protein expression is generally known to those of skilled in the art of molecular biology, for example, see Sambrook et al., Molecular cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. , 1989. The promoter employed may be inducible. The term "inducible" used in the context of a promoter means that the promoter only directs transcription of an operably linked nucleotide sequence if a stimulus is present, such as a change in temperature or the presence of a chemical substance ("chemical inducer"). As used herein, "chemical induction" according to the present invention refers to the physical application of a exogenous or endogenous substance (incl. macromolecules, e.g., proteins or nucleic acids) to a host cell. This has the effect of causing the target promoter present in the host cell to increase the rate of transcription. Alternatively, the promoter employed may be constitutive. The term "constitutive" used in the context of a promoter means that the promoter is capable of directing transcription of an operably linked nucleotide sequence in the absence of stimulus (such as heat shock, chemicals etc.).

Non-limiting examples of promoters functional in bacteria, such as Bacillus subtilis, Lactococcus lactis or Escherichia coli, include both constitutive and inducible promoters such as T7 promoter, the beta-lactamase and lactose promoter systems; alkaline phosphatase (phoA) promoter, a tryptophan (trp) promoter system, tetracycline promoter, lambda-phage promoter, ribosomal protein promoters; and hybrid promoters such as the tac promoter. Other bacterial and synthetic promoters are also suitable.

Non-limiting examples of promoters functional in yeast, such as *Saccharomyces cerevisiae*, include xylose promoter, GAL1 and GAL10 promoters, TEF1 promoter, and pgk1 promoter.

Non-limiting examples of promoters functional in fungi, such as *Aspergillus Oryzae* or *Aspergillus niger*, include promotors derived from the gene encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral α-amylase, *Aspergillus niger* acid stable α-amylase, *Aspergillus niger* or *Aspergillus awamsii* glucoamylase (gluA), *Aspergillus niger* acetamidase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphatase isomerase, *Rhizopus meihei* aspartic proteinase, and *Rhizopus meihei* lipase.

Non-limiting examples of promoters functional in alga, such as *Haematococcus pluvialis*, include the CaMV35S promoter, the SV40 promoter, and promoter of the *Chlamydomonas reinhardtii* RBCS2 gene and the promoter of the *Volvox carteri* ARS gene.

Non-limiting examples of promoters functional in plant cells include the Lactuca sative psbA promoter, the tabacco psbA promoter, the tobacco rrn16 PEP+NEP promoter, the CaMV 35S promoter, the 19S promoter, the tomate E8 promoter, the nos promoter, the Mac promoter, and the pet E promoter or the ACT1 promoter.

Besides a promoter, the exogenous nucleic acid molecule may further comprise at least one regulatory element selected from a 5' untranslated region (5'UTR) and 3' untranslated region (3' UTR). Many such 5' UTRs and 3' UTRs derived from prokaryotes and eukaryotes are well known to the skilled person. Such regulatory elements include 5' UTRs and 3' UTRs normally associated with other genes, and/or 5' UTRs and 3' UTRs isolated from any bacteria, yeast, fungi, alga or plant cell.

If the host cell is a prokaryotic organism, the 5' UTR usually contains a ribosome binding site (RBS), also known as the Shine Dalgarno sequence which is usually 3-10 base pairs upstream from the initiation codon. Meanwhile, if the host cell is an eukaryotic organism the 5' UTR usually contains the Kozak consensus sequence. An eukaryotic 5' UTR may also contain cis-acting regulatory elements.

The exogenous nucleic acid molecule may be a vector or part of a vector, such as an expression vector. Normally, such a vector remains extrachromosomal within the host cell which means that it is found outside of the nucleus or nucleoid region of the host cell.

It is also contemplated by the present invention that the exogenous nucleic acid molecule is stably integrated into the genome of the host cell. Means for stable integration into the genome of a host cell, e.g., by homologous recombination, are well known to the skilled person.

It is understood that the details given herein with respect to a recombinant host cell apply to other aspects of the invention, in particular to the processes according to the invention, which are described in more detail below.

Methods and uses

The present invention provides processes for the production of sulfated phenolic compounds. Particularly, a process for the production of a sulfated phenolic compound is provided comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification ; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

According to certain embodiments, the process for the production of a sulfated phenolic compound comprises:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

According to other certain embodiments, the process for the production of a sulfated phenolic compound comprises:

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host

cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

According to other certain embodiments, the process for the production of a sulfated phenolic compound comprises:

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

The medium employed may be any conventional medium suitable for culturing the host cell in question, and may be composed according to the principles of the prior art. The medium will usually contain all nutrients necessary for the growth and survival of the respective host cell, such as carbon and nitrogen sources and other inorganic salts. Suitable media, e.g. minimal or complex media, are available from commercial suppliers, or may be prepared according to published receipts, e.g. the American Type Culture Collection (ATCC) Catalogue of strains. Non-limiting standard medium well known to the skilled person include Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, MS broth, Yeast Peptone Dextrose, BMMY, GMMY, or Yeast Malt Extract (YM) broth, which are all commercially available. A non-limiting example of suitable media for culturing bacterial cells, such as B. subtilis, L. lactis or E. coli cells, including minimal media and rich media such as Luria Broth (LB), M9 media, M17 media, SA media, MOPS media, Terrific Broth, YT and others. Suitable media for culturing eukaryotic cells, such as yeast cells, are RPMI 1640, MEM, DMEM, all of which may be supplemented with serum and/or growth factors as required by the particular host cell being cultured. The medium for culturing eukaryotic cells may also be any kind of minimal media such as Yeast minimal media.

The fermentable carbon substrate may be any suitable carbon substrate know in the art, and in particularly any carbon substrate commonly used in the cultivation of microorganisms and/ or fermentation. Non-limiting examples of suitable fermentable carbon substates include carbohydrates (e.g., C5 sugars such as arabinose or xylose, or C6 sugars such as glucose), glycerol, glycerine, acetate, dihydroxyacetone, one-carbon source, methanol, methane, oils, animal fats, animal oils, plant oils, fatty acids, lipids,

phospholipids, glycerolipids, monoglycerides, diglycerides, triglycerides, renewable carbon sources, polypeptides (e.g., a microbial or plant protein or peptide), yeast extract, component from a yeast extract, peptone, casaminoacids or any combination of two or more of the foregoing.

According to certain embodiments, the carbon substate is selected from the group consisting of C5 sugars (such as arabinose or xylose), C6 sugars (such as glucose or fructose), lactose, sucrose, glycerol, glycerine, acetate, Corn steep liquor, yeast extract, component from a yeast extract, peptone, casaminoacids or combinations thereof.

According to certain embodiments, the medium comprises glucose.

According to certain other embodiments, the medium comprises glycerol.

According to certain other embodiments, the medium comprises acetate.

It is also contemplated to use starch as a carbon substrate. Depending on the microorganism used, the metabolization of starch may require the supplementation of beta-glucosidase, such as the beta-glucosidase from Neurospora crassa, to the medium. Alternatively, a recombination host cell according to the invention may be further genetically modified to express a beta-glucosidase, such as the beta-glucosidase from *Neurospora crassa*.

When a fermentable carbon substrate is employed it is thus possible that the recombinant host cell produces the phenolic compound or a precursor thereof directly from such primary carbon substrate.

Therefore, according to certain embodiments, the process for the production of a sulfated phenolic compound comprises: (i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell, a phenolic compound being produced from the fermentable carbon substrate by the first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

According to certain embodiments, the process further comprises:

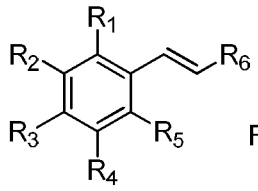
(ii) culturing the first recombinant host cell under suitable conditions for the production of the corresponding sulfated phenolic compound.

Suitable conditions for culturing the respective host cell are well known to the skilled person. Typically, the recombinant host cell is cultured at a temperature ranging from about 23 to about 60°C, such as from about 25 to about 40°C, such as at about 37°C. The pH of the medium may range from pH 1.0 to pH 14.0, such as from about pH 1 to about pH 2, from about pH 4 to about pH 11, from about pH 5 to about pH 10, from about pH 6 to about pH 10, or from about pH 7 to about pH 9.5, e.g. at pH 6.0, pH pH 7.0, pH. 7.5, pH 8.0, pH 8.5, pH 9.0, pH 9.5, pH 10.0, pH 10.5 or pH 11.0.

The process may further comprise iii) recovering the sulfated phenolic compound. The sulfated phenolic compound may be recovered by conventional method for isolation and purification chemical compounds from a medium. Well-known purification procedures include centrifugation or filtration, precipitation, and chromatographic methods such as e.g. ion exchange chromatography, gel filtration chromatography, etc.

For the purpose of this specification and the appended claims, it should be understood that the phenolic compounds include those compounds in which a hydroxyl group is directly attached to a benzenoid carbon atom, and which compounds may or may not contain other substituent groups.

According to certain embodiments, the phenolic compound is a compound represented by the general formula (I):



Formula (I)

;

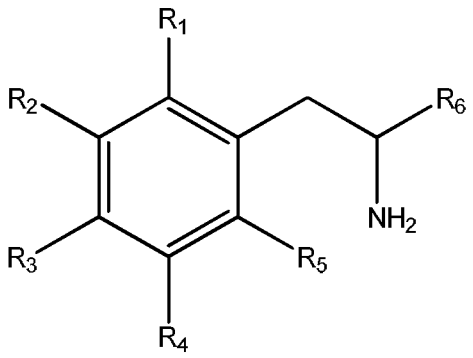
wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

Specific examples of compounds of Formula I include, but are not limited to, reservatrol, o-, m-, and p-coumaric acid, caffeic acid, ferulic acid, sinapic acid, curcumin, rosmarinic acid, sinapyl alcohol, coniferyl alcohol, and salvianolic acid.

A precursor of a phenolic compound according to Formula I may be a compound represented by the general Formula (p-I):



Formula (p-I);

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

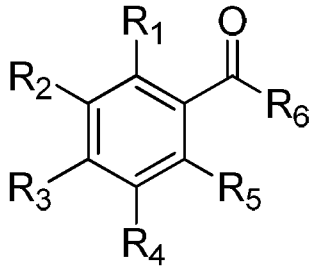
wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

Such a precursor may be converted to the phenolic compound by a recombinant host cell according to the invention, comprising a polypeptide having tyrosine ammonia lyase activity. Such polypeptide will eliminate ammonia from the precursor of Formula (p-I) under the formation of the corresponding molecule of Formula I. Preferably, the p-I precursor is the L-isomer.

Accoring to certain embodiments, the precursor of a phenolic compound as employed in step (i''') is a compound of the general Formula (p-I) as defined herein.

According to certain other embodiments, the phenolic compound is a compound represented by the general formula (II):



Formula (II)

;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

According to certain embodiments, R_6 is $-COOR_7$.

According to certain embodiments, R_7 is hydrogen.

According to certain embodiments, R_2 is hydroxyl (-OH).

According to certain embodiments, R_3 is hydroxyl (-OH).

According to certain embodiments, R_4 is hydroxyl (-OH).

According to certain embodiments, each of R_1, R_2, R_4 and R_5 is hydrogen.

According to certain embdodiments, each of R_1, R_2, and R_5 is hydrogen.

According to particular embodiments, the phenolic compound is p-coumaric acid (Formula I: R_1=H, R_2=H, R_3=OH, R_4=H, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic compound is caffeic acid (Formula I: R_1=H, R_2=H, R_3=OH, R_4=OH, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic acid is ferulic acid (Formula I: R_1=H, R_2=OCH_3, R_3=OH, R_4=H, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic acid is isoferulic acid (Formula I: R1=H, R2=OH, R3= OCH_3, R4=H, R5= H, R6=COOH).

According to other particular embodiments, the phenolic acid is sinapic acid (Formula I: R_1=H, R_2=OCH_3, R_3=OH, R_4=OCH_3, R_5= H, R_6=COOH).

According to other particular embodiments, the phenolic compound is resveratrol (Formula I: R_1=H, R_2=OH, R_3=H, R_4=OH, R_5= H, R_6=p-hydroxyphenyl).

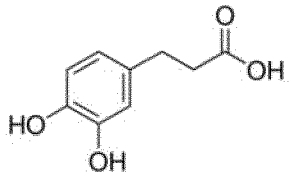
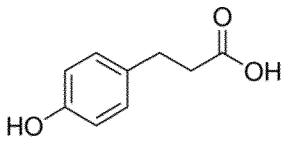
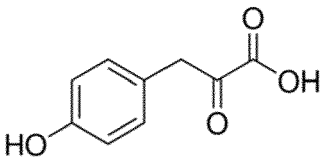
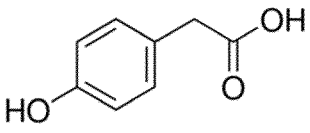
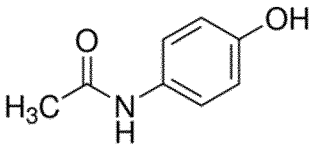
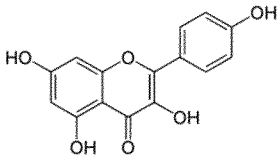
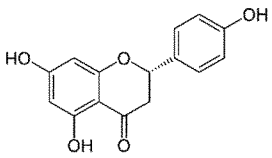
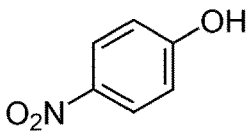
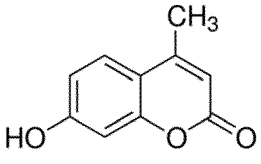
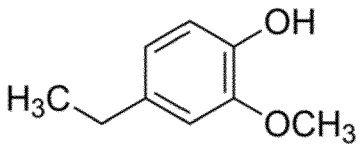
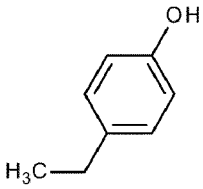
According to other particular embodiments, the phenolic compound is vanillin (Formula II: R_1=H, R_2=H, R_3=OH, R_4= OCH_3, R_5= H, R_6=H).

According to other particular embodiments, the phenolic compound is vanillic acid (Formula II: R1=H, R2=H, R3=OH, R4= OCH_3, R5= H, R6=OH).

According to other particular embodiments, the phenolic compound is 4-vinylphenol (Formula I: R_1=H, R_2=H, R_3=OH, R_4=H, R_5=H, R_6=H).

According to other particular embodiments, the phenolic compound is 2-methoxy 4-vinylphenol (Formula I: R_1=H, R_2= OCH_3, R_3=OH, R_4=H, R_5=H, R_6=H).

According to certain embodiments, the phenolic compound is selected from the group consisting of:


(3,4-Dihydroxyhydrocinnamic acid),  (3-(4-Hydroxyphenyl)propionic acid),  (4-hydroxyphenylpyruvic acid),  (4-Hydroxyphenylacetic acid),  (4-acetamidophenol),  (Kaempferol),  (Naringenin),  (4-nitrophenol),  (4-methylbelliferone),  (4-ethylguaiacol), and  (4-ethylphenol).

According to particular embodiments, the phenolic compound is 3,4-Dihydroxyhydrocinnamic acid.

According to other particular embodiments, the phenolic compound is 3-(4-Hydroxyphenyl)propionic acid.

According to other particular embodiments, the phenolic compound is 4-hydroxyphenylpyruvic acid.

According to other particular embodiments, the phenolic compound is 4-Hydroxyphenylacetic acid.

According to other particular embodiments, the phenolic compound is 4-acetamidophenol.

According to other particular embodiments, the phenolic compound is Kaempferol.

According to other particular embodiments, the phenolic compound is Naringenin.

According to other particular embodiments, the phenolic compound is 4-nitrophenol.

According to other particular embodiments, the phenolic compound is 4-methylbelliferone.

According to other particular embodiments, the phenolic compound is 4-ethylguaiacol.

According to other particular embodiments, the phenolic compound is 4-ethylphenol.

According to other particular embodiments, the phenolic compound is Luteolin.

According to other particular embodiments, the phenolic compound is Apigenin.

According to other particular embodiments, the phenolic compound is fisetin.

According to other particular embodiments, the phenolic compound is Quercetin.

According to certain embodiments, the phenolic compound is a hydroxycinnamic acid.

According to certain embodiments, the phenolic compound is a compound represented by the general formula (I), wherein R_1 is hydrogen; R_2, R_3 and R_4 independently are selected from the group consisting of hydrogen (H), hydroxyl (-OH), C_{1-6}-alkyl and C_{1-6}-Alkoxy, provided that at least one of R_2, R_3 and R_4 is hydroxyl (-OH); R_5 is hydrogen, and R_6 is COOH.

According to certain embodiments, the precursor of a phenolic compound as employed in step (i''') is a compound of the general Formula (p-I), wherein R_1 is hydrogen; R_2, R_3 and R_4 independently are selected from the group consisting of hydrogen (H), hydroxyl (-OH), C_{1-6}-alkyl and C_{1-6}-Alkoxy, provided that at least one of R_2, R_3 and R_4 is hydroxyl (-OH); R_5 is hydrogen, and R_6 is COOH.

According to certain embodiment, the sulfated phenolic compound obtained in according to the present invention is zosteric acid.

Suitable sulfate donor molecules metabolized by a polypeptide having aryl sulfotransferase activity are well-known to one skilled in the art. Non-limiting examples include 3'-phosphoadenosine 5'-phosphosulfate (PAPS), para-nitrophenyl sulfate (pNPS) and 4-methylumbelliferyl sulfate (MUS). Such sulfate donor molecules may be employed to facilitate the sulfation of phenolic compounds in accordance with the invention.

The medium employed for culturing the recombinant host cell may be any conventional medium suitable for culturing the host cell in question, and may be composed according to the principles of the prior art. The medium will usually contain all nutrients necessary for the growth and survival of the respective host cell, such as carbon and nitrogen sources and other inorganic salts, such as sulfate salts. Suitable media, e.g. minimal or complex media, are available from commercial suppliers, or may be prepared according to published receipts, e.g. the American Type Culture Collection (ATCC) Catalogue of strains. Non-limiting standard medium well known to the skilled person include Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth, MS broth, Yeast Peptone Dextrose, BMMY, GMMY, or Yeast Malt Extract (YM) broth, which are all commercially available. A non-limiting example of suitable media for culturing bacterial cells, such as B. subtilis, L. lactis or E. coli cells, including minimal media and rich media such as Luria Broth (LB), M9 media, M17 media, SA media, MOPS media, Terrific Broth, YT and others. Suitable media for culturing eukaryotic cells, such as yeast cells, are RPMI 1640, MEM, DMEM, all of which may be supplemented with serum and/or growth factors as required by the particular host cell being cultured. The medium for culturing eukaryotic cells may also be any kind of minimal media such as Yeast minimal media.

The invention described and claimed herein is not to be limited in scope by the specific embodiments herein disclosed, since these embodiments are intended as illustrations of several aspects of the invention. Any equivalent embodiments are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the invention.

Certain definitions

"Aryl sulfotransferase activity" as used herein refers to the ability of a polypeptide to catalyze the catalyze the transfer of a sulfate group from a donor molecule to an aryl acceptor molecule.

"Tyrosine ammonia lyase activity" as used herein refers to the ability of a polypeptide to catalysed the conversion of L-tyrosine into p-coumaric acid.

"Phenylalanine ammonia lyase activity" as used herein refers to the ability of a polypeptide to catalysed the conversion of L-phenylalanine into trans-cinnamic acid.

"Sulfate transporter" or "sulfate permease" are used herein interchangeably to refer to a protein or protein complex that mediates sulfate uptake by a cell.

"ATP sulfurylase" as used herein refers to an enzyme that catalyzes the reaction: ATP + sulfate = diphosphate + adenosine 5'-phosphosulfate (APS).

"APS kinase" as used herein refers to an enzyme that catalyzes the reaction: ATP + adenosine 5'-phosphosulfate (APS) = ADP + 3'-phosphoadenosine 5'-phosphosulfate (PAPS).

"PAP phosphatase" as used herein refers to an enzyme that catalyzes the reaction: 3'-phosphoadenosine 5'-phosphate (PAP)+ H_2O = AMP + phosphate.

"Polypeptide," or "protein" are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-transiational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids.

"Nucleic acid" or "polynucleotide" are used interchangeably herein to denote a polymer of at least two nucleic acid monomer units or bases (e.g., adenine, cytosine, guanine, thymine) covalently linked by a phosphodiester bond, regardless of length or base modification.

"Recombinant" or "non-naturally occurring" when used with reference to, e.g., a host cell, nucleic acid, or polypeptide, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature, or is identical thereto but produced or derived from synthetic

materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant host cells expressing genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise expressed at a different level.

"Substitution" or "substituted" refers to modification of the polypeptide by replacing one amino acid residue with another, for instance the replacement of an Arginine residue with a Glutamine residue in a polypeptide sequence is an amino acid substitution.

"Conservative substitution" refers to a substitution of an amino acid residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain may be substituted with another aliphatic amino acid, e.g., alanine, valine, leucine, and isoleucine; an amino acid with hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain, e.g., serine and threonine; an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain, e.g., phenylalanine, tyrosine, tryptophan, and histidine; an amino acid with a basic side chain is substituted with another amino acid with a basic side chain, e.g., lysine and arginine; an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain, e.g., aspartic acid or glutamic acid; and a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

"Non-conservative substitution" refers to substitution of an amino acid in a polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

"Deletion" or “deleted” refers to modification of the polypeptide by removal of one or more amino acids in the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the polypeptide while retaining enzymatic activity and/or retaining the improved properties of an engineered enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide, in various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

"Insertion" or “inserted” refers to modification of the polypeptide by addition of one or more amino acids to the reference polypeptide. Insertions can comprise addition of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the reference polypeptide.

"Host cell" as used herein refers to a living cell or microorganism that is capable of reproducing its genetic material and along with it recombinant genetic material that has been introduced into it - e.g., via heterologous transformation.

"Expression" includes any step involved in the production of a polypeptide (e.g., encoded enzyme) including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

As used herein, "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid molecule to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded nucleic acid loop into which additional nucleic acid segments can be ligated. Certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". Certain other vectors are capable of facilitating the insertion of a exogenous nucleic acid molecule into a genome of a host cell. Such vectors are referred to herein as "transformation vectors". In general, vectors of utility in recombinant nucleic acid

techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of a vector. Large numbers of suitable vectors are known to those of skill in the art and commercially available.

As used herein, "promoter" refers to a sequence of DNA, usually upstream (5') of the coding region of a structural gene, which controls the expression of the coding region by providing recognition and binding sites for RNA polymerase and other factors which may be required for initiation of transcription. The selection of the promoter will depend upon the nucleic acid sequence of interest. A "promoter functional in a host cell" refers to a "promoter" which is capable of supporting the initiation of transcription in said cell, causing the production of an mRNA molecule.

As used herein, "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence. A promoter sequence is "operably-linked" to a gene when it is in sufficient proximity to the transcription start site of a gene to regulate transcription of the gene.

As used herein, an operon is a functioning unit of DNA containing a cluster of genes under the control of a single promoter.

"Percentage of sequence identity," "% sequence identity" and "percent identity" are used herein to refer to comparisons between an amino acid sequence and a reference amino acid sequence. The “% sequence identify”, as used herein, is calculated from the two amino acid sequences as follows: The sequences are aligned using Version 9 of the Genetic Computing Group's GAP (global alignment program), using the default BLOSUM62 matrix (see below) with a gap open penalty of -12 (for the first null of a gap) and a gap extension penalty of -4 (for each additional null in the gap). After alignment, percentage identity is calculated by expressing the number of matches as a percentage of the number of amino acids in the reference amino acid sequence.

The following BLOSUM62 matrix is used:

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
Ala	4																			
Arg	-1	5																		
Asn	-2	0	6																	
Asp	-2	-2	1	6																
Cys	0	-3	-3	-3	9															
Gln	-1	1	0	0	-3	5														
Glu	-1	0	0	2	-4	2	5													
Gly	0	-2	0	-1	-3	-2	-2	6												
His	-2	0	1	-1	-3	0	0	-2	8											
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

"Reference sequence" or “reference amino acid sequence” refers to a defined sequence to which another sequence is compared. In the context of the present invention a reference amino acid sequence may be any amino acid sequence set forth in SEQ ID NO: 1 to 50.

Aliphatic radicals/groups, as referred herein, are optionally mono-or polysubstituted and may be branched or unbranched, saturated or unsaturated. Unsaturated aliphatic groups, as defined in herein, include alkyl, alkenyl and alkinyl radicals. Preferred aliphatic radicals according to the present invention include but are not restricted to methyl, ethyl, vinyl (ethenyl), ethinyl, propyl, n-propyl, isopropyl, allyl (2-propenyl), 1-propinyl, methylethyl, butyl, n-butyl, iso-butyl, sec-butyl, tert-butyl butenyl, butinyl, 1-methylpropyl, 2-methylpropyl, 1,1-dimethylethyl, pentyl, n-pentyl, 1,1-dimethylpropyl, 1,2-dimethylpropyl, 2,2-dimethylpropyl, hexyl, 1-methylpentyl, n-heptyl, n-octyl, n-nonyl and n-decyl. Preferred substituents for aliphatic radicals, according to the present invention, are a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, oxo, (C=O)R', SR', SOR', SO_2R', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents a linear or branched C_{1-6}-alkyl group.

“Alkyl”, “alkyl radical” or group as used herein means saturated, linear or branched hydrocarbons, which can be unsubstituted or mono- or polysubstituted. Thus, unsaturated alkyl is understood to encompass alkenyl and alkinyl groups, like e.g. -CH=CH-CH_3 or -C≡C-CH_3, while saturated alkyl encompasses e.g. -CH_3 and -CH_2-CH_3. “C_{1-12}-alkyl” includes C_{1-2}-

alkyl, C_{1-3}-alkyl, C_{1-4}-alkyl, and C_{1-5}-alkyl, C_{1-6}-alkyl, C_{1-7}-alkyl, C_{1-8}-alkyl, C_{1-9}-alkyl, C_{1-10}-alkyl, and C_{1-11}-alkyl. In these radicals, C_{1-2}-alkyl represents C_1- or C_2-alkyl, C_{1-3}-alkyl represents C_1-, C_2- or C_3-alkyl, C_{1-4}-alkyl represents C_1-, C_2-, C_3- or C_4-alkyl, C_{1-5}-alkyl represents C_1-, C_2-, C_3-, C_4-, or C_5-alkyl, C_{1-6}-alkyl represents C_1-, C_2-, C_3-, C_4-, C_5- or C_6-alkyl etc. The alkyl radicals may be methyl, ethyl, vinyl (ethenyl), propyl, allyl (2-propenyl), 1-propinyl, methylethyl, butyl, 1-methylpropyl, 2-methylpropyl, 1,1-dimethylethyl, pentyl, 1,1-dimethylpropyl, 1,2-dimethylpropyl, 2,2-dimethylpropyl, hexyl, 1-methylpentyl, if substituted also CHF_2, CF_3 or CH_2OH etc. These alkyl, alkenyl or alkinyl radicals may optionally be mono-or polysubstituted by substitutents independently selected from a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, (C=O)R', SR', SOR', SO_2R', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents linear or branched C_{1-6}-alkyl group.

"Aryl" or "aryl radical" as herein is understood as meaning ring systems with at least one aromatic ring but without heteroatoms even in only one of the rings. These aryl radicals may optionally be mono-or polysubstituted by substitutents independently selected from a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, an optionally at least mono-substituted phenyl group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, oxo, (C=O)R', SR', SOR', SO_2R', N(C=O)-OR', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents a linear or branched C_{1-6}-alkyl group. Preferred examples of aryl radicals include but are not restricted to phenyl, naphthyl, fluoranthenyl, fluorenyl, tetralinyl or indanyl or anthracenyl radicals, which may optionally be mono- or polysubstituted, if not defined otherwise.

"Alkyl-aryl" or "alkyl-aryl radical" as used herein comprises a linear or branched, optionally at least mono-substituted alkyl chain which is bonded to an aryl group, as defined above. A preferred alkyl-aryl radical is a benzyl group, wherein the alkyl chain is optionally branched or substituted. Preferred substituents for alky-aryl radicals, according to the present invention, are F, Cl, Br, I, NH_2, SH, OH, SO_2, CF_3, carboxy, amido, cyano, carbamyl, nitro, phenyl, benzyl, $-SO_2NH_2$, C_{1-6} alkyl and/or C_{1-6}-alkoxy.

"Heteroaryl" or "heteroaryl radical" as used herein is understood as meaning heterocyclic ring systems which have at least one aromatic ring and may optionally contain one or more heteroatoms from the group consisting of nitrogen, oxygen and/or sulfur and may

optionally be mono-or polysubstituted by substitutents independently selected from a C_{1-4} alkyl group, a linear or branched C_{1-6} alkoxy group, F, Cl, I, Br, CF_3, CH_2F, CHF_2, CN, OH, SH, NH_2, oxo, (C=O)R', SR', SOR', SO_2R', NHR', NR'R" whereby R' and optionally R" for each substitutent independently represents a linear or branched C_1-6-alkyl group. Preferred examples of heteroaryls include but are not restricted to furan, benzofuran, thiophene, benzothiophene, pyrrole, pyridine, pyrimidine, pyridazine, pyrazine, quinoline, isoquinoline, phthalazine, benzo-1,2,5-thiadiazole, benzothiazole, indole, benzotriazole, benzodioxolane, benzodioxane, benzimidzole, carbazole and quinazoline.

"Alkoxy", "alkoxy radical" or group as used herein means an "alkyl" singular bonded to oxygen. "C_{1-6}-alkoxy" includes C_{1-2}- alkoxy, C_{1-3}-alkoxy, C_{1-4}-alkoxy, and C_{1-5}-alkoxy, as well as C_{2-3}-alkoxy, C_{2-4}-alkoxy, C_{2-5}-alkoxy, C_{3-4}-alkoxy, C_{3-5}-alkoxy, and C_{4-5}-alkoxy. In these radicals, C_{1-2}- alkoxy represents C1- or C2- alkoxy, C_{1-3}- alkoxy represents C_1-, C_2- or C_3-alkoxy, C_{1-4}-alkyl represents C_1-, C_2-, C_3- or C_4-alkoxy, C_{1-5}-alkoxy represents C_1-, C_2-, C_3-, C_4-, or C_5-alkoxy, C_{1-6}- alkoxy represents C_1-, C_2-, C_3-, C_4-, C_5- or C_6-alkoxy. The alkoxy radicals may be methoxy, ethoxy, propoxy, butoxy, pentyloxy or hexyloxy.

The term "precursor of a phenolic compound" refers to any compound that may be converted to a phenolic compound by a host cells as described herein.

Where a numerical limit or range is stated herein, the endpoints are included. Also, all values and sub ranges within a numerical limit or range are specifically included as if explicitly written out.

Having generally described this invention, a further understanding can be obtained by reference to certain specific examples, which are provided herein for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

Examples

Example 1 – Production of zosteric acid in *E. coli*

A range of aryl sulfotransferases including SULT1A1 *Rattus norvegicus* (SEQ ID NO: 1), SULT1A1 *Homo sapiens* (SEQ ID NO: 2), SULT1A1 *Equus caballus* (SEQ ID NO: 3), SULT1A1 *Sus scrofa domesticus* (SEQ ID NO: 4), SULT1A1 *Canis lupus familiaris* (SEQ ID NO: 5) and SULT1E1 *Gallus gallus domesticus* (SEQ ID NO: 6) were expressed in *Escherichia coli*. The respective genes encoding SEQ ID NO. 1, 3, 4, 5, and 6 were cloned amplified from liver

tissue cDNA (Zyagen) by PCR using the primers listed in Table 1. The nucleotide sequence of the gene encoding SEQ ID NO: 2 was codon optimized for expression in *Escherichia coli* (GeneArt, Life Technologies) and amplified by PCR using the primers in Table 1. The pETDuet-1 plasmid was digested with restriction endonucleases *Nco*I and *Sal*I. The PCR products were then individually cloned into the plasmid pETDuet-1 using the Gibson reaction (New England Biolabs). The resulting plasmids were transformed into BL21(DE3)pLysS (Life Technologies). Figure 1 shows the plasmid map of the plasmid encoding SULT1A1 *Rattus norvegicus* (SEQ ID NO: 1).

Table 1: Overview of enzymes and primers for cloning aryl sulfotransferases

SEQ ID NO	Name	Fwd Primer	Rev Primer
1	SULT1A1 Rattus norvegicus	CBJP472	CBJP473
2	SULT1A1 Homo sapiens	CBJP470	CBJP471
3	SULT1A1 Equus caballus	CBJP499	CBJP500
4	SULT1A1 Sus scrofa domesticus	CBJP505	CBJP506
5	SULT1A1 Canis lupus familiaris	CBJP503	CBJP504
6	SULT1E1 Gallus gallus domesticus	CBJP501	CBJP502

The strains were grown in M9 minimal media containing glucose as a carbon source, and 0.1 mM IPTG for induction of gene expression as well as 0.1 mM *p*-coumaric acid (pHCA). After four days of growth, samples were withdrawn by filtration and analyzed by HPLC.

The concentration of p-courmaric acid (pHCA) and zosteric acid in the supernatant was quantified by high performance liquid chromatography (HPLC) and compared to chemical standards. HPLC was done on a Thermo setup using a HS-F5 column and mobile phases: 5 mM ammonium formate pH 4.0 (A) and acetonitrile (B) at 1.5 mL min-1, using a gradient elution starting at 5% B. From 0.5 min after injection to 7 min, the fraction of B increased linearly from 5% to 60%, and between 9.5 min and 9.6 the fraction of B decreased back to 5%, and remaining there until 12 min. pHCA and zosteric acid were quantified by measuring absorbance at 277 nm.

Table 2 shows the remaining pHCA and the produced zosteric acid in the culture media. Zosteric acid was formed with an aryl sulfotransferase heterologously expressed in a microorganism exemplified by *E. coli* supplied with the substrate.

Table 2: Production of zosteric acid in E. coli from pHCA through the heterologous expression of sulfotransferases.

Enzyme	pHCA remaining (mM)	Zosteric acid formed (mM)
No enzyme	0.10	Not detectable
SULT1A1 Rattus norvegicus	0.02	0.10
SULT1A1 Homo sapiens	0.08	0.02
SULT1A1 Equus caballus	0.09	0.01
SULT1A1 Sus scrofa domesticus	0.09	0.01
SULT1A1 Canis lupus familiaris	0.10	0.01
SULT1E1 Gallus gallus domesticus	0.08	0.01

Example 2 – Increased production of zosteric acid in *E. coli*

The addition of sulfated groups to targets is **dependent** on supply of the donor molecule 3'-Phosphoadenosine 5'-phosphosulfate (PAPS). We examined if we could increase the production of zosteric acid by overexpressing enzymes providing PAPS and an enzyme that removes the product 3'-Phosphoadenosine 5'-phosphate (PAP).

Table 3: Cloning of enzymes involved in activating sulfate and product removal.

Genes	Fwd Primer	Rev Primer
cysDNC alone	CBJP491	CBJP492
cysDNC for artificial operon	CBJP491	CBJP497
cysQ for artificial operon	CBJP498	CBJP496

In E. coli, the genes cysD and cysN encode the two subunits of ATP sulfurylase (EC:2.7.7.4), cysC encodes APS kinase (EC:2.7.1.25), and cysQ encode a PAP phosphatase.

The cysDNC cluster was amplified by PCR from *E. coli* MG1655 chromosomal DNA using the primers shown in Table3. The plasmid pRSFDuet-1 (Life Technologies) was digested by the restriction endonucleases NdeI and BglII. The gene cluster was inserted into the digested plasmid using the Gibson reaction (New England Biolabs). Figure 2 shows the resulting plasmid. For the combined expression of cysDNC and cysQ in an artificial operon, cysDNCQ, the two parts were amplified by PCR from *E. coli* MG1655 chromosomal DNA using the primers shown in Table 3. Again the parts were inserted into the digested plasmid. Figure 3 shows the resulting plasmids. The plasmid expressing SULT1A1 Homo sapiens (SEQ ID NO: 2) from example 1 was co-transformed into *E. coli* BL21(DE3)pLysS cells (Life Technologies) with either the plasmid expressing cysDNC or cysDNCQ.

Cells were grown as in Example 1 and the supernatants were analyzed for product formation as in example 1. The strain expressing SULT1A1 in combination with cysDNCQ

was also grown without the addition of IPTIG for induction. Table 4 shows the concentrations of pHCA and zosteric acid.

Table 4: Concentrations of pHCA and zosteric acid in culture media with E. coli expressing an aryl sulfotransferase in combination with cysDNC and cysQ.

Enzymes	Induction	pHCA remaining (mM)	Zosteric acid formed (mM)
SULT1A1 Homo sapiens	0.1 mM IPTG	0.08	0.02
SULT1A1 Homo sapiens, CysDNC	0.1 mM IPTG	0.06	0.06
SULT1A1 Homo sapiens, CysDNCQ	0.1 mM IPTG	0.04	0.09
SULT1A1 Homo sapiens, CysDNCQ	None	0.10	Not detectable

This shows that more of the pHCA is transformed into zosteric acid when the protein expression of cysDNC is increased. Even more zosteric acid is formed when the protein expression cysQ is additionally increased.

Example 3 – A sulfated product can be formed in vivo by co-expression of an heterologous pathway and an aryl sulfotransferase

The production of a sulfated product can be accomplished biologically by the expression of aryl sulfotransferase as shown in example 1. The substrate for sulfation may also be formed by a biological organism, and here it will be shown for an organism expressing both a heterologous pathway leading to a phenolic compound and expressing a sulfotransferase acting upon the phenolic compound.

The enzyme RmXAL from Rhodotorula mucilaginosa / Rhodotorula rubra (SEQ ID NO: 46) has tyrosine ammonia lyase activity, thus catalyzing the non-oxidative deamination of the amino acid tyrosine, releasing p-coumaric acid (pHCA) and ammonia. The gene encoding RmXAL was codon optimized using standard algorithms for expression in *E. coli* available by GeneArt (Life Technologies) and amplified by PCR using the primers shown in Table 5 and inserted into the pCDFDuet-1 vector (Novagen / Life Technologies), which had been digested by the restriction enzymes NdeI and BglII, using Gibson reaction (New England Biolabs).

Table 5: Primers used for cloning of tyrosine ammonialyase

Genes	Fwd Primer	Rev Primer
RmXAL	CBJP487	CBJP488

The resulting plasmid was co-transformed into *E. coli* BL21(DE3)pLysS cells (Life Technologies) alone or together with the plasmid expressing SULT1A1 from Homo sapiens (example 1). The resulting strains was grown in M9 media with glucose as a carbon source, with 0.1 mM IPTG for induction of gene expression. Samples were taken as described previously (example 1) for analysis of product formation. Table 6 shows the resulting concentrations of pHCA and zosteric acid. RmXAL allowed the production of pHCA without addition of any substrate, thus providing a heterologous pathway from the cells normal metabolism to a heterologous product. The additional expression of an aryl sulfotransferase, exemplified by SULT1A1 from *Homo sapiens*, allowed the *in vivo* conversion of pHCA to zosteric acid. Thus, an aryl sulfotransferase can act upon a compound produced *in vivo* and the cells can release the resulting sulfated product to the medium.

Table 6: Concentrations of pHCA and zosteric acid in culture media with E. coli expressing an aryl sulfotransferase in combination with a tyrosine ammonia lyase.

Enzymes	pHCA (mM)	Zosteric acid formed (mM)
RmXAL	0.04	Not detectable
SULT1A1 Homo sapiens, RmXAL	0.02	0.01

Example 4 – Decreased toxicity of sulfated product

E. coli MG1655 was grown in chemically defined M9 minimal media with 0.2% glucose as a carbon source without further addition or with the additions of either 10 mM, 20 mM, 25 mM, 30 mM, 35 mM or 40 mM p-coumaric acid (pHCA), or with 20 mM or 40 mM of the sulfate ester of pHCA (zosteric acid). All media preparations had been adjusted to pH 7. Cells were grown at 37°C with 250 rpm shaking in an orbital shaker. The growth rates were examined by following the optical density at 600 nm. The resulting growth rates in exponential growth phase are shown in Figure 4. Filled squares represent growth rates in media with pHCA. Open squares represent growth rates in media with zosteric acid. And the circle represents the growth rate in media without any of these additions. It is evident

that the presence of pHCA is toxic to the cells, while the sulfate ester, zosteric acid is much less so.

Example 5 – In vivo supply of precursor of sulfated product

The substrate that is the subject for sulfation may be supplied to the medium void of such precursors or may be provided by microorganisms in the medium. Here we show that *p*-coumaric acid that is sulfated to generate zosteric acid, can be produced in vivo by the expression of a tyrosine ammonia-lyase.

The genes encoding the tyrosine ammonia-lyases RcTAL (from *Rhodobacter capsulatus*; SEQ ID NO: 50), RsTAL (from *Rhodobacter sphaeroides*; SEQ ID NO: 43) and FjTAL (from *Flavobacterium johnsoniae*; SEQ ID NO: 40) were cloned into expression vectors as follows. Genes (SEQ ID NO: 51, 52, and 53, respectively) were optimized for *E. coli* and synthesized by GeneArt, amplified by PCR using the oligonucleotides shown in the table below, and cloned into pCDFDuet-1 (Novagen): The plasmid was digested with *Nde*I and *Bgl*II and gel purified. The genes were inserted by isothermal assembly using Gibson Assembly Master Mix (New England Biolabs), and transformed into chemically competent DH5α (laboratory strain) or NEB5α (New England Biolabs), selecting for resistance to 50 μg mL^{-1} spectinomycin in LB medium. Resulting plasmids pCBJ215 (RsTAL), pCBJ228 (FjTAL) and pCBJ297 (RcTAL) were co-transformed by electroporation into the *E. coli* expression strain BL21(DE3) (Invitrogen/Life Technologies) together with a pETDuet-1-based plasmid expressing SULT1A1 from rat (Example 1). Transformation cultures were plated on LB containing 50 μg mL^{-1} spectinomycin and 100 μg mL^{-1} ampicillin. A control strain carrying pCDFDuet-1 was also made.

Table 7: Primers

Oligonucleotide	Gene	Direction	Sequence
CBJP483	RsTAL	Forward	CATCTTAGTATATTAGTTAAGTATAAGAAGGAGATATACATATGCTGGCAATGAGCCCT
CBJP484	RsTAL	Reverse	TGGCCGGCCGATATCCAATTGATTAAACCGGACTCTGTTGC
CBJP555	FjTAL	Forward	CATCTTAGTATATTAGTTAAGTATAAGAAGGAGATATACATATGAACACCATCAACGAATATCTG
CBJP556	FjTAL	Reverse	TGGCCGGCCGATATCCAATTGATTAATTGTTAATCAGGTGGTCTTTTACTTTCTG
CBJP745	RcTAL	Forward	CATCTTAGTATATTAGTTAAGTATAAGAAGGAGATATACATATGCTGGATGCAACCATTGG

CBJP746	RcTAL	Reverse	TGGCCGGCCGATATCCAATTGATTATGCCGGAGGA TCCGCT

Strains harboring recombinant plasmids were pre-cultured in 2xYT liquid medium with 100 µg mL^{-1} ampicillin and 50 µg mL^{-1} spectinomycin and incubated at 37°C and 250 rpm overnight. The following day, each pre-culture was transferred into 5 ml of M9 minimal medium with 0.2% glucose, 2 mM tyrosine and 1 mM IPTG for induction of expression. Cultures were placed in an incubator at 37°C with shaking at 250 rpm overnight. The supernatants were then collected by centrifugation twice and applied to HPLC analysis as described in example 1, and the titers of *p*-coumaric acid (pHCA) and zosteric acid (ZA) were quantified using chemical standards and are presented in the table below.

Table 8: Titers of p-coumaric acid (pHCA) and zosteric acid (ZA)

Sulfotransferase	Tyrosine ammonia-lyase	µM pHCA	µM ZA
SULT1A1 rat	None	0	0
SULT1A1 rat	RsTAL	78	<1
SULT1A1 rat	RcTAL	20	<1
SULT1A1 rat	FjTAL	398	16

Here, it is evident that the zosteric acid is formed when there is a supply of exogenous *p*-coumaric acid or if the cells are capable of producing *p*-coumaric acid. Conclusively, a sulfated product may be formed from an unsulfated precursor molecule, when this is produced in vivo.

Furthermore, the data surprisingly show that employing the tyrosine ammonia-lyase from *Flavobacterium johnsoniae* (FjTAL; SEQ ID NO: 40) results in a higher supply in unsulfated precursor molecule (here: p-coumaric acid), which in turn leads to a higher yield of sulfated product (here: zosteric acid) compare to other tyrosine ammonia-lyases.

Example 6 – Production of sulfated products in other hosts

We have shown that zosteric acid can be produced in vivo in *Escherichia coli* by expression of an aryl sulfotransferase. To show that the reaction is possible in other microorganisms, we here show that the yeast *Saccharomyces cerevisiae* can also be used as a host for the production.

The gene encoding aryl sulfotransferase SULT1A (Example 1) was cloned after a TEF1 promoter into an episomal plasmid with a 2-micron origin of replication as follows. The gene was amplified by PCR using primers CBJP633 and CBJP634. Alternatively, the gene was codon-optimized for *E. coli* and synthesized by GeneArt and amplified by primers CBJP635 and CBJP636. The TEF1 promoter (Jensen et al., 2014, *FEMS Yeast Res* 14: 238-248) was amplified by PCR using the primers PTEF1_fw and PTEF1_rv. Plasmid pCfB132 (Jensen et al., supra) was digested by restriction enzymes AsiSI and Nt.BsmI. The three fragments – plasmid, TEF1 promotor and SULT1A1-encoding gene – were assembled using a uracil-excision cloning procedure, resulting in plasmids pCBJ283 and pCBJ284, which were subsequently transformed into the *Saccharomyces cerevisiae* strain CEN.PK102-5B selecting for growth on synthetic dropout media plates lacking uracil. A control strain was also made by transformation of pCfB132 into CEN.PK102-5B.

Table 9: Primers

Oligonucleotide	Gene/promoter	Direction	Sequence
CBJP633	SULT1A1 rat	Forward	AGTGCAGGUAAAACAATGgagttctcccgtcca
CBJP634	SULT1A1 rat	Reverse	CGTGCGAUTCAtagttcacaacgaaacttg
CBJP635	SULT1A1 rat (E. coli)	Forward	ATCTGTCAUAAAACAATGgaattttcacgtccgc
CBJP636	SULT1A1 rat (E. coli)	Reverse	CACGCGAUTCAcagttcacaacgaaatttgaa
PTEF1_fw	PTEF1	Forward	Cacgcgaugcacacaccatagcttc
PTEF1_rv	PTEF1	Reverse	Cgtgcgauggaagtaccttcaaaga

The strains were grown in modified Delft medium (Jensen et al., supra) with 20 mg/mL histidine and 60 mg/mL leucine and 10 mM *p*-coumaric acid overnight at 30°C with aeration. The supernatant was then isolated and examined by HPLC as described in Example 1. The table below shows that zosteric acid (ZA) was produced by the strain expressing SULT1A1 and not the control strain lacking a sulfotransferase.

Table 10: Titers of zosteric acid

Sulfotransferase	**µM ZA** (averages and standard deviations of replicate experiments)
None	0 ± 0
SULT1A1 rat (native)	37.8 ± 5.7
SULT1A1 rat (codon optimized for E. coli)	46.2 ± 3.5

It is evident that zosteric acid is formed only when a sulfotransferase is expressed in yeast, and that the gene encoding this may be natural or encoded by a synthetic gene with a specific codon-optimization. Conclusively, the sulfation reactions shown to be catalyzed by sulfotransferases in *E. coli* are also catalyzed when the sulfotransferases are expressed in other organisms, as demonstrated here for the yeast *S. cerevisiae*. The efficacy of production may be affected by means such as the codon-usage of the genes encoding the sulfotransferase. Thus yeast expressing sulfotransferases may be able to detoxify aromatic compounds such as *p*-coumaric acid, and form sulfated products such as zosteric acid.

Example 7 – A range of compounds are substrates for sulfation in vivo

Here we show that the expression of an aryl sulfotransferase may be able to convert several substrates. Some of these are inhibitors that can be found in biomass hydrolyzate used as a substrate for cell growth and production in biotechnology. The compounds also include some that are of biotechnological interest as products of a cell culture or be some whose sulfate ester is of economic interest.

Different sulfotransferases were examined for their substrate specificities against three substrates. We tested the sulfotransferases mentioned in example 1, as well as additional ones. The genes encoding these were cloned as described in example 1 using the primers shown in the table below from cDNA libraries of the respective organisms, except for the SULT1A1 from rat (*Rattus norvegicus*) codon-optimized for *E. coli* (described above). The resulting vectors were transformed into BL21(DE3)pLysS.

Table 11: Primers

Oligonucleotide	Gene	Direction	
CBJP517	SULT1C1 *Gallus gallus domesticus*	Forward	TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCatggccctggataaaatgg
CBJP518	SULT1C1 *Gallus gallus domesticus*	Reverse	TAAGCATTATGCGGCCGCAAGCTTGtcacaattccatgcgaaaaactag
CBJP533	SULT1A1 *Rattus norvegicus* (Codon-optimized for E. coli)	Forward	TAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCatggaattttcacgtcc
CBJP534	SULT1A1 *Rattus norvegicus* (Codon-optimized for E. coli)	Reverse	TAAGCATTATGCGGCCGCAAGCTTGttacagttcacaacgaaatttg

The resulting strains were grown in M9 medium containing either 100 µM pHCA, 95 µM resveratrol or 87 µM kaempferol. The cultures were grown overnight at 37°C, 300 rpm. The following day the supernatants were isolated and examined by HPLC as described in example 1. BL21(DE3)pLysS were used as a control strain and did not convert the substrates.

Table 12: Percent conversion of the various substrates

Enzyme	**pHCA**	**resveratrol**	**kaempferol**
	100 µM	95 µM	87 µM
SULT1A1 *Rattus norvegicus*	93%	93%	95%
SULT1C1 *Gallus gallus domesticus*	26%	100%	80%
SULT1A1 *Rattus norvegicus* (Codon-optimized for *E. coli*)	73%	58%	38%
SULT1A1 human	39%	36%	97%
SULT1A1 *Equus caballus*	21%	100%	96%
SULT1E1 *Gallus gallus domesticus*	17%	100%	47%
SULT1A1 *Canis lupus familiaris*	34%	61%	60%
SULT1A1 *Sus scrofa domesticus*	8%	88%	45%

The table shows the percent conversion of the various substrates by cells expressing the different sulfotransferases. The results show that several sulfotransferases, and especially the aryl sulfotransferase from rat (*Rattus norvegicus*), may be employed in the sulfation of phenolic compounds.

To further test the range of substrates that can be sulfated, we used strains carrying plasmids expressing SULT1A1 from rat (*Rattus norvegicus*) and SULT1E1 from chicken (*Gallus gallus domesticus*) (Example 1) cloned into the expression vector pETDuet-1, and cysDNCQ from *E. coli* cloned into expression vector pRSFDuet-1 (Example 2). The plasmids

were introduced into the *E. coli* expression strain BL21(DE3)pLysS as described previously, selecting for transformants with appropriate antibiotics, namely 34 µg mL^{-1} chloramphenicol for pLysS, 100 µg mL^{-1} ampicillin for pETDuet-1-based vectors, and 100 µg mL^{-1} kanamycin for pRSFDuet-1-based vectors. The table below shows the combination of over-expressed genes on plasmids. A control strain without a sulfotransferase gene or cysDNCQ operon was also examined.

Table 13: Combination of over-expressed genes on plasmids

E. coli strains	Sulfotransferase	Cys genes
Control strain	-	-
SULT1A1 rat	SULT1A1 rat	-
SULT1E1 chicken	SULT1E1 chicken	-
SULT1A1 rat + CysDNCQ	SULT1A1 rat	CysDNCQ

The strains were precultured in 2xYT medium with appropriate antibiotics. 10 µL of these precultures were used to inoculate M9 media with 1 mM IPTG and none or a single substrate for sulfation. After overnight growth at 37°C, 300 rpm the supernatants were withdrawn and examined by HPLC as described in Example 1. The compounds were detected by UV absorbance. The table below shows the percent reduction in concentration in the strains expressing sulfotransferases alone or in combination with cysDNCQ genes when compared to the control strain.

Table 14: Percent reduction in concentration

Compound	Start concentration in µM	SULT1A1	SULT1E1	SULT1A1 + CysDNCQ
Ferulic acid	110	72%	67%	100%
Quercetin	85	75%	74%	81%
4-hydroxybenzoic acid	287	5%	4%	6%
4-acetamidophenol	114	24%	10%	30%
3-Hydroxy-4-methoxycinnamic acid	132	51%	24%	62%
4-Hydroxyphenylpyruvic acid	255	47%	100%	64%

3-(4-Hydroxyphenyl)propionic acid	241	3%	1%	7%
Vanillic acid	173	33%	0%	39%
Luteolin	61	27%	0%	37%
Apigenin	77	41%	98%	99%
fisetin	81	98%	98%	100%

Conclusively, a wide range of phenolic compounds are substrates for sulfotransferases. In the shown examples, the conversion is enhanced by the overexpression of cysDNCQ genes. Some of these compounds and their sulfate esters are of interest in biotechnology. Also, some of these compounds are inhibitors of cell growth and function, and thus conversion by sulfation is of interest for use in biological systems.

Example 8 – Increasing uptake of sulfate

E. coli BL21(DE3)-derived strains expressing an aryl sulfotransferase and a sulfate transporter were constructed as follows.

A plasmid (Figure 5) was constructed for the over-expressing of the CysZ (NCBI reference sequence NP_416908.1) from *E. coli* by amplifying the cysZ gene from the chromosome of E. coli MG1655 by PCR using the primers in the table below. The resulting PCR product and the plasmid expressing SULT1A1 from rat mentioned in example 1 (Figure 1) were digested using the restriction enzymes *HindIII* and *NotI*, purified by column purification and ligated together with T4 DNA polymerase, and the ligation reaction was used to transform the *E. coli* cloning strain NEB5α (New England Biolabs). Resulting colonies resistant to ampicillin were tested for correct insert by PCR using primers pET-Upstream (ATGCGTCCGGCGTAGA) and DuetDOWN1 (GATTATGCGGCCGTGTACAA). The correct plasmid was purified and transformed into *E. coli* BL21(DE3) (Life Technologies) together with the plasmid encoding CysDNCQ from example 2 (Figure 3), selecting for both ampicillin and kanamycin.

Similarly a plasmid (Figure 6) was constructed for the over-expressing of the CysP (GenBank AAC75478.1), CysT (GenBank AAC75477.1, CysW (GenBank AAC75476.2) and CysA (Genbank AAC75475.1) from *E. coli*. The cysPTWA (also known as cysPUWA) operon was amplified from the chromosome of E. coli MG1655 by PCR using the primers in the table

below. The resulting PCR product and the plasmid expressing CysDNCQ mentioned in example 2 (Figure 3) were digested using restriction enzymes HindIII and NotI, purified by column purification and ligated together with T4 DNA polymerase, and the ligation reaction was used to transform the *E. coli* cloning strain NEB5α. Resulting colonies resistant to kanamycin were tested for correct insert by PCR using primers ACYCDuetUP1 (GGATCTCGACGCTCTCCCT) and DuetDOWN1 (GATTATGCGGCCGTGTACAA). The correct plasmid was purified and transformed into *E. coli* BL21(DE3) together with the plamid encoding SULT1A1 from rat mentioned in example 1 (Figure 1), selecting for both ampicillin and kanamycin.

Table 15: Primers

Oligonucleotide	Gene(s)	Direction	Sequence	Restriction site
CBJP891	*cysZ*	Forward	ttaaaagcttgggattggtcaaaaggagctcatcc	*Hin*dIII
CBJP892	*cysZ*	Reverse	aatagcggccgcttaccgccacatcgcgtgtttat	*Not*I
CBJP893	*cysPTWA*	Forward	ttaaaagcttagaaagtcattaaatttataagggtgcgca	*Hin*dIII
CBJP894	*cysPTWA*	Reverse	aatagcggccgctcaggcgctttgtgcgagagc	*Not*I

Control strains carrying only the emptive plasmids pETDuet-1 and pRSFDuet-1 (Life Technologies), carrying the plasmid encoding SULT1A1 from rat (Figure 1) and pRSFDuet-1, or carrying the plasmid encoding SULT1A1 from rat (Figure 1) and the plasmid encoding cysDNCQ (Figure 3) were used as controls for growth experiments.

The *E. coli* strains were propagated overnight in grown in M9 minimal medium containing 0.2 % (w/v) glucose, 2 mM p-coumaric acid, 100 µg/mL ampicillin and 50 µg/mL kanamycin in wells of a 96-well deep-well plate (EnzyScreen) shaking at 300 rpm in an orbital shaker at 37°C. From these cultures, 30 µL was used to inoculate 500 µL of M9 medium with 0.2% (w/v) glucose, 2 mM p-coumaric acid, 100 µg/mL ampicillin, 50 µg/mL kanamycin and further 200 µM IPTG in wells of a 96-well deep-well plate (EnzyScreen) that was left shaking at 300 rpm in an orbital shaker at 37°C overnight. The cell density was then measured by the optical density at 600 nm, and the supernatant was sampled for production of zosteric acid, by two rounds of centrifugation. Zosteric acid in the supernatant was quantified by high performance liquid chromatography (HPLC) and compared to a chemical standard. HPLC was done on a Thermo setup using a HS-F5 column (3 µm) and mobile phases: 5 mM

ammonium formate pH 4.0 (A) and acetonitrile (B) at 1.5 mL min-1, using a gradient elution starting at 5% B. From 0.5 min after injection to 7 min, the fraction of B increased linearly from 5% to 60%, and between 9.5 min and 9.6 the fraction of B decreased back to 5%, and remaining there until 12 min. Zosteric acid were quantified by measuring absorbance at 290 nm.

Figure 7 shows that increased titers of zosteric acid were reached when overexpressing CysZ and CysPTWA.

It is clear that the additional expression of a sulfate transporter such as that encoded by *cysZ* (proton symporter) or by *cysPTWA* (ABC transporter) enhanced the sulfation catalyzed by a phenol sulfotransferase, optionally with increased expression of sulfate adenylyltransferase, APS kinase and adenosine-3',5'-bisphosphate nucleotidase. The sulfate transport activity may be optained by the activity of transporters belonging to several different families of transporters.

Example 9 – Different compounds as substrates

Further compounds may be sulfated by sulfotransferases and cells heterologously expressing such, which is shown by the following experiment.

E. coli strain KRX (obtained from Promega) was transformed with either plasmids pETDuet-1 or the derived plasmid encoding SULT1A1 from *Rattus norvegicus* described in Example 1.

M9 medium containing 0.2% glucose, 0.1 mM IPTG, 0.1 % rhamnose, and 100 µg/mL ampicillin was prepared. To aliquots of the medium, phenolic compounds (Table 14) were added from 10 mM stock solutions in 99.9 % ethanol to a final concentration of the compounds of 50 µM.

The strains described above were grown in 2xYT medium with 100 µg/mL ampicillin overnight before their were used to inoculate the media by 50-fold dilution. The cultures were grown overnight with vigorous shaking at 37 °C. The supernatants were isolated by centrifugation and subjected to HPLC analysis as described in example 1. The compounds were detected by UV absorbance except for 4-methylbelliferone, which was measured by fluorescence.

Table 16 shows that the phenolic compounds were all subject to sulfation, when a sulfotransferase is present in the medium. Additional peaks corresponding to more

hydrophilic compounds were also the result of the activity of the sulfotransferase on each of the compounds.

Table 16: Percent reduction

Compound	SULT1A1
Sinapic acid	31%
Naringenin	60%
4-ethylphenol	100%
4-vinylphenol	100%
4-ethylguaiacol	50%
4-methylbelliferone	100%
4-nitrophenol	30%

Conclusively, a wide range of phenolic compounds are substrates for sulfotransferases. Some of these compounds and their sulfate esters are of interest in biotechnology or as markers of sulfation activity or as donors of sulfate in the reverse reaction. Also, some of these compounds are inhibitors of cell growth and function, and thus conversion by sulfation is of interest for use in biological systems. The example shows that the phenolic acceptor molecule for sulfation may differ by the position of the hydroxyl-group and still remains an active substrate. 4-vinylphenol is a degradation product of *p*-coumaric acid by decarboxylation, and it is still a substrate for sulfation by a sulfotransferase, showing that the side chain can vary significantly, and the compound remains an active substrate.

Example 10 - Different sulfotransferases are active

The sulfotransferases may be of very different sequences. To show this, we tested the sulfotransferases with lower homology to the sequences presented in the examples above against three substrates.

The gene (sequence SEQ ID NO: 98) encoding dmST1 (sequence SEQ ID NO: 99) from *Drosophila melanogaster* was amplified from cDNA using primers listed in Table 17 similarly to the cloning of genes described in Example 1.

Table 17: Primers

Oligonucleotide	Gene	Direction	Sequence
CBJP474	dmST1 *Drosophila melanogaster*	Forward	TAGAAATAATTTTGTTTAACTTTAA GAAGGAGATATAC C ATGCCCCAGTCGAGCTTCTT

CBJP475	dmST1 *Drosophila melanogaster*	Reverse	TAAGCATTATGCGGCCGCAAGCTTG TTACGTGGACGCAAACTTGCT

The gene (sequence SEQ ID NO: 100) encoding SULT1ST1 (sequence SEQ ID NO: 101) from *Danio rerio* was codon-optimized for *E. coli* and synthesized as in example 1.

The gene (sequence SEQ ID NO: 102) encoding SULT6B1 (sequence SEQ ID NO: 103) from *Danio rerio* was codon-optimized for *E. coli* and synthesized as in example 1.

The gene (sequence SEQ ID NO: 104) encoding Hoch_6098 (sequence SEQ ID NO: 105) from the bacterium *Haliangium ochraceum* DSM 14365 was codon-optimized for E. coli and synthesized as in example 1.

The genes were cloned into the vector pETDuet-1 as described in Example 1, and the resulting vectors were transformed into *E. coli* KRX (obtained from Promega). Strains were grown as described above in presence of either 100 µM resveratrol, 20 µM kaempferol or 50 µL 3-hydroxy-4-methoxycinnamic acid, and the supernatant were analyzed as described above, except that for kaempferol the cultures were mixed with an equal volume of methanol before isolation of the supernatants. Table 18 shows the reduction in the concentrations of compounds in the presence sulfotransferases (n.d. = not determined).

Table 18: Percent reduction

	Resveratrol	Kaempferol	3-hydroxy-4-methoxycinnamic acid
D. rerio SULT1ST1	80%	n.d.	12%
D. rerio SULT6B1	n.d.	44%	n.d.
D. melanogaster dmST1	n.d.	n.d.	12%
H. ochraceum DSM 14365 Hoch_6098	7%	100%	n.d.

The example shows that the sulfation reaction may occur in a medium with a cell expressing a heterologous sulfotransferase more distantly related to the sequences in the previous examples. It may even be of non-animal origin, exemplified with the bacterial sulfotransferase from *Haliangium ochraceum*.

Claims

1. A process for the production of a sulfated phenolic compound comprising:

(i') contacting a medium comprising a phenolic compound with a first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i'') contacting a medium comprising a fermentable carbon substrate with a first recombinant host cell, a phenolic compound being produced from the fermentable carbon substrate by the first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification; or

(i''') contacting a medium comprising a precursor of a phenolic compound with a first recombinant host cell, the precursor being converted to the phenolic compound by the first recombinant host cell; wherein the first recombinant host cell comprises a heterologous polypeptide having an aryl sulfotransferase activity, and wherein the first recombinant host cell has been modified to have an increased uptake of sulfate compared to an identical host cell that does not carry said modification.

2. The process according to item 1, further comprising:

(ii) culturing the first recombinant host cell under suitable conditions for the production of the corresponding sulfated phenolic compound; and

(iii) optionally, recovering said sulfated phenolic compound.

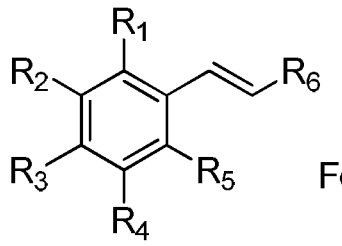
3. The process according to claim 1 or 2, wherein the heterologous polypeptide having an aryl sulfotransferase activity is selected from the group consisting of:

1a) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13;

1b) a polypeptide comprising an amino acid sequence which has at least about 70% sequence identity to the amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein said polypeptide has aryl sulfotransferase activity; or

1c) a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 or 13, wherein 1 to 50 amino acid residues are substituted, deleted and/or inserted, and wherein said polypeptide has aryl sulfotransferase activity.

4. The process according to any one of claims 1-3, wherein the phenolic compound is represented by the general formula (I):



Formula (I)

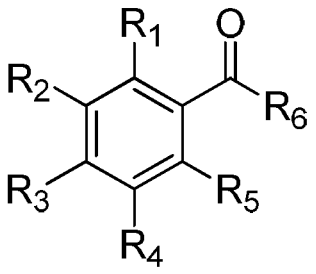
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

5. A process according to any one of the claims 1-3, wherein the phenolic compound is represented by the general formula (II):



Formula (II)

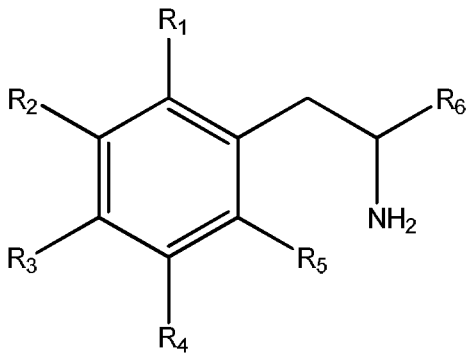
;

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, $-OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

6. The process according to any one of claims 1-3, wherein the precursor of a phenolic compound in step (i''') is a compound of the general Formula (p-I):



Formula (p-I);

wherein at least one of R_1, R_2, R_3, R_4, and R_5 being an hydroxyl group (-OH);

wherein R_1, R_2, R_3, R_4, R_5 and R_6 are independently selected from the group consisting of halide, hydrogen, hydroxyl (-OH), $-OR_7$, $-OCOR_7$, $-NR_7R_8$, $-COR_7$, $-COOR_7$, $-SR_7$, $-OSO_3R_7$, -

$OCSR_7$, $-POR_7R_8$, alkyl, alkenyl, alkynyl, aryl, and heteroaryl; wherein R_7, and R_8 are independently selected from the group consisting of hydrogen, alkyl, alkenyl, alkynyl, aryl, and heteroaryl;

wherein R_1, R_2, R_3, R_4, R_5 and R_6, are optionally linked with a bridge member Y_n, thereby forming one or more rings, Y_n being a bond or a C_{1-12} alkyl or an aryl, a carbocyclic, a heterocyclic or a heteroaromatic structure having 1-3 rings, 3-8 ring members in each and 0 to 4 heteroatoms, or a heteroalkyl comprising 1 to 12 heteroatoms selected from the group consisting of N, O, S, $S(O)_{1-2}$ and carbonyl, and wherein n is an integer between 1 and 12.

7. The process according to any one of claims 1-6, wherein said first recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

8. The process according to claim 7, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter, by modifying the ribosome binding site and/or by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

9. The process according to any one of claims 7 or 8, wherein the sulfate transporter is a selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

10. The process according to any one of claims 1 to 9, wherein the first recombinant host cell has further been modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification; wherein the recombinant host cell has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification; and/or wherein the recombinant host cell has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

11. A recombinant host cell comprising a heterologous polypeptide having aryl sulfotransferase activity; wherein the recombinant host cell has been modified to have an

increased uptake of sulfate compared to an identical host cell that does not carry said modification.

12. The recombinant host cell according to claim 11, wherein said recombinant host cell has been modified to an have increased protein expression of a sulfate transporter compared to the identical host cell that does not carry said modification.

13. The recombinant host cell according to claim 12, wherein the increase in protein expression of the sulfate transporter is achieved by increasing the number of copies of a gene or genes encoding said sulfate transporter, by modifying the ribosome binding site and/or by increasing the strength of the promoter(s) operably linked to the gene or genes encoding said sulfate transporter.

14. The recombinant host cell according to any one of claims 12 or 13, wherein the sulfate transporter is a selected from the group consisting of: members of the CysZ family, members of the SulT (cysPTWA) family, members of the SulP family, CysP transporters belonging to the phosphate inorganic transporter (PiT) family, and oxyanion permeases (PerO).

15. The recombinant host cell according to any one of claims 11 to 14, wherein the recombinant host cell has further been modified to have an increased protein expression of an ATP sulfurylase compared to an identical host cell that does not carry said modification; wherein the recombinant host cell has further been modified to have an increased protein expression of an APS kinase compared to an identical host cell that does not carry said modification; and/or wherein the recombinant host cell has further been modified to have an increased protein expression of a PAP phosphatase compared to an identical host cell that does not carry said modification.

16. Use of the recombinant host cell according to any one of claims 11-15 in the production of a sulfated phenolic compound.

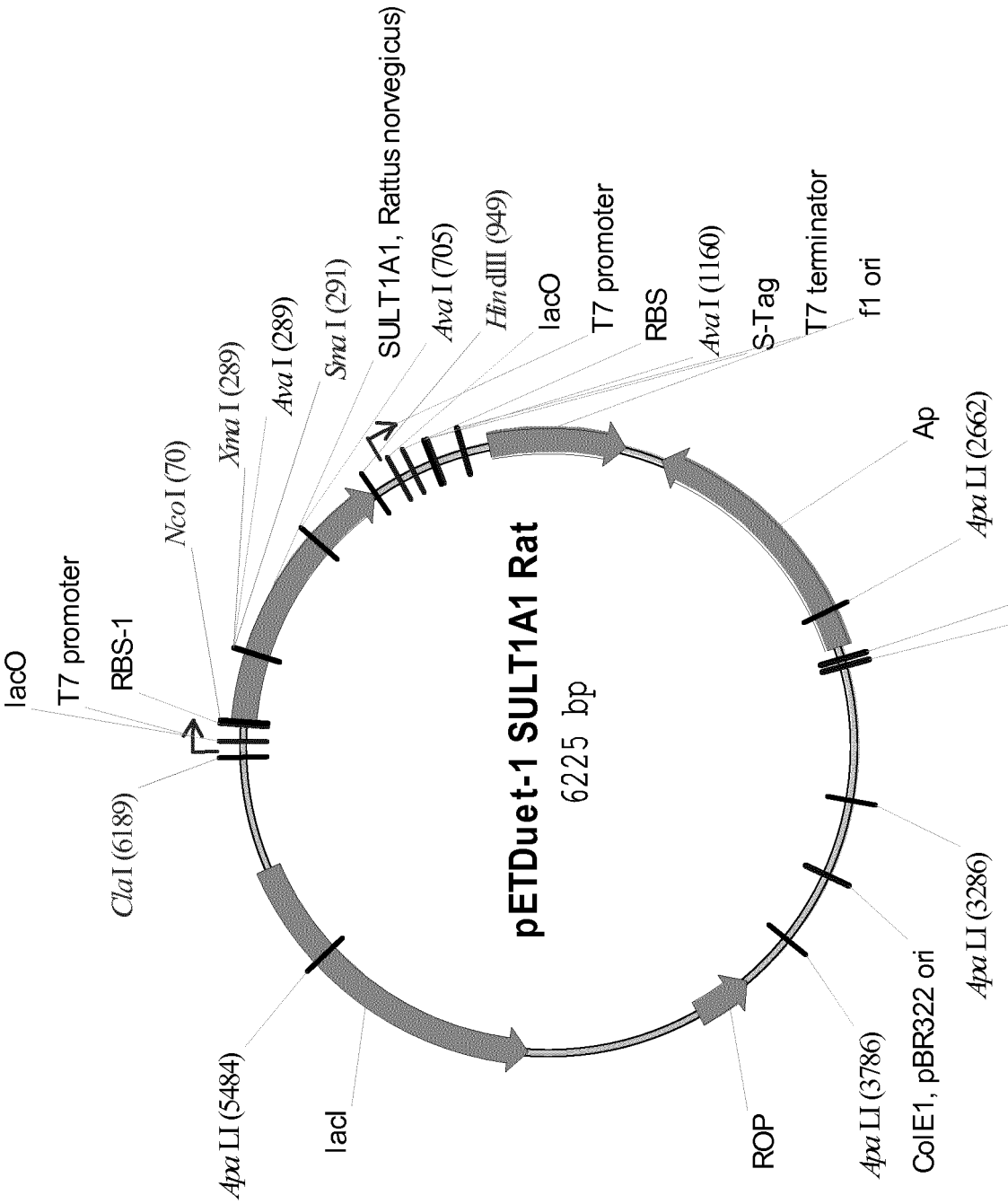
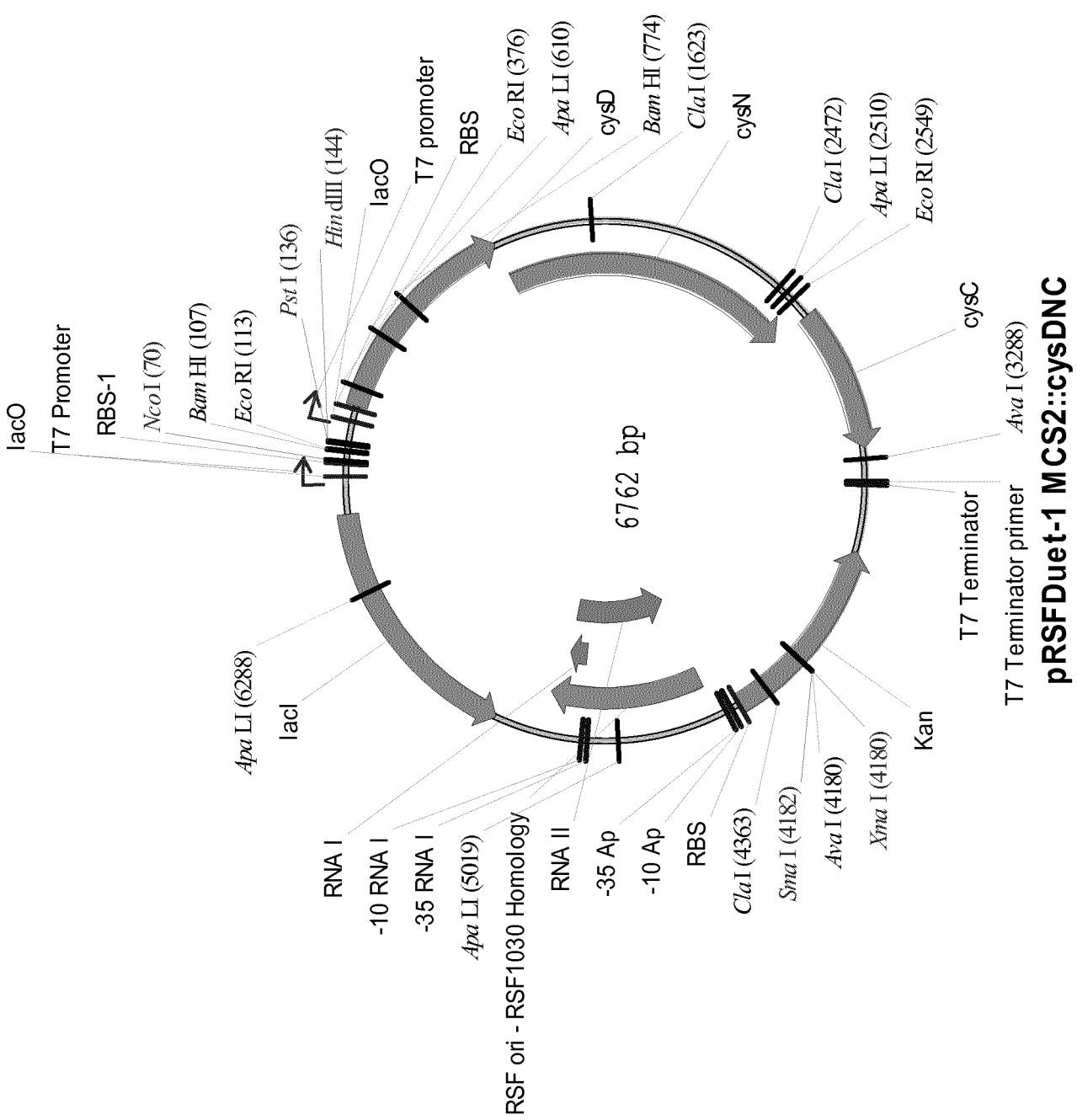
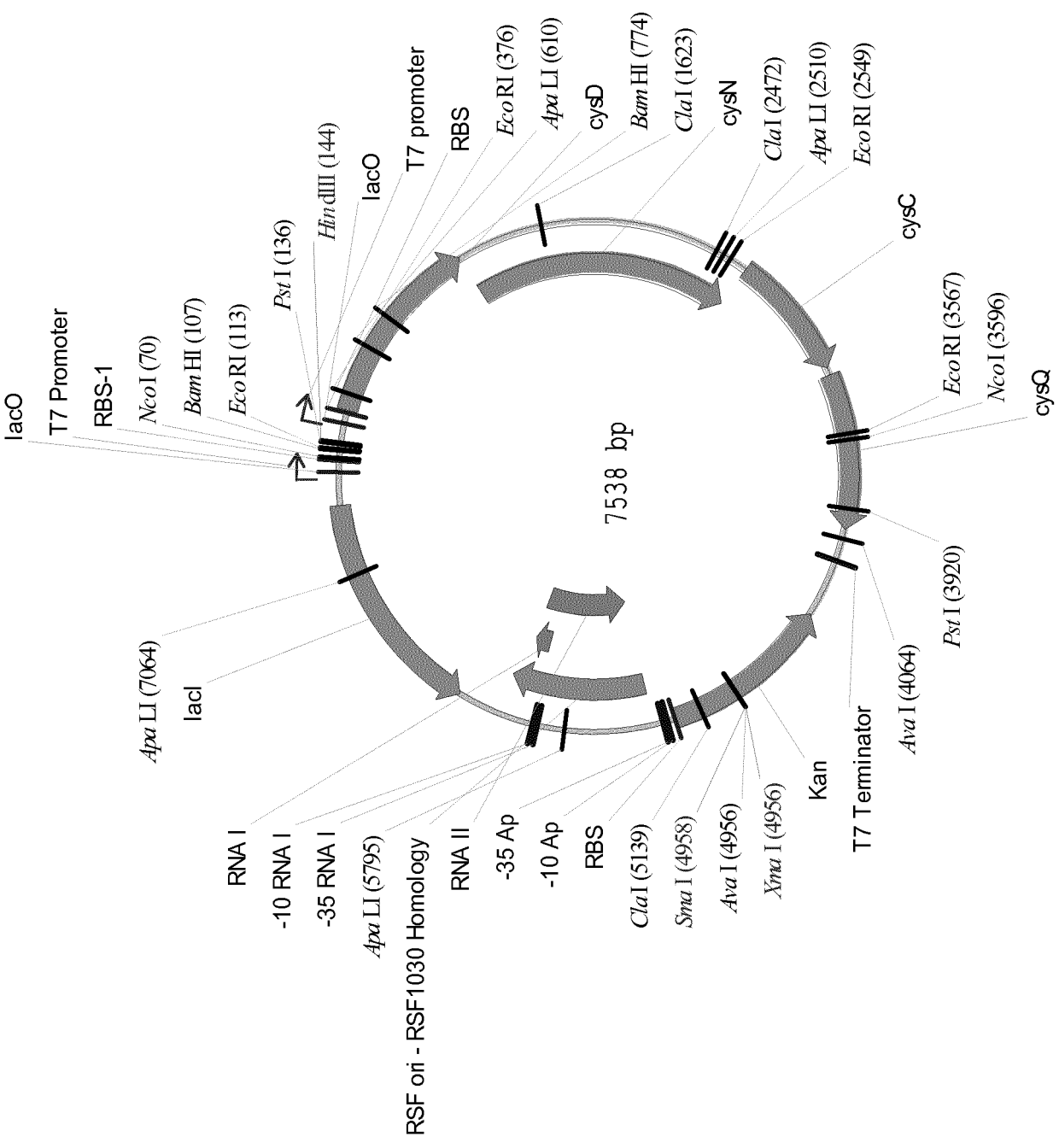

pETDuet-1 SULT1A1 Rat
6225 bp
lacO
T7 promoter
RBS-1
NcoI (70)
XmaI (289)
AvaI (289)
SmaI (291)
SULT1A1, Rattus norvegicus)
AvaI (705)
HindIII (949)
lacO
T7 promoter
RBS
AvaI (1160)
S-Tag
T7 terminator
f1 ori
Ap
ApaLI (2662)
-10 Ap
-35 Ap
ApaLI (3286)
ColE1, pBR322 ori
ApaLI (3786)
ROP
lacI
ApaLI (5484)
ClaI (6189)

Figure 1



pRSFDuet-1 MCS2::cysDNC

Figure 2



pRSFDuet-1 MCS2::cysDNC-Q

Figure 3

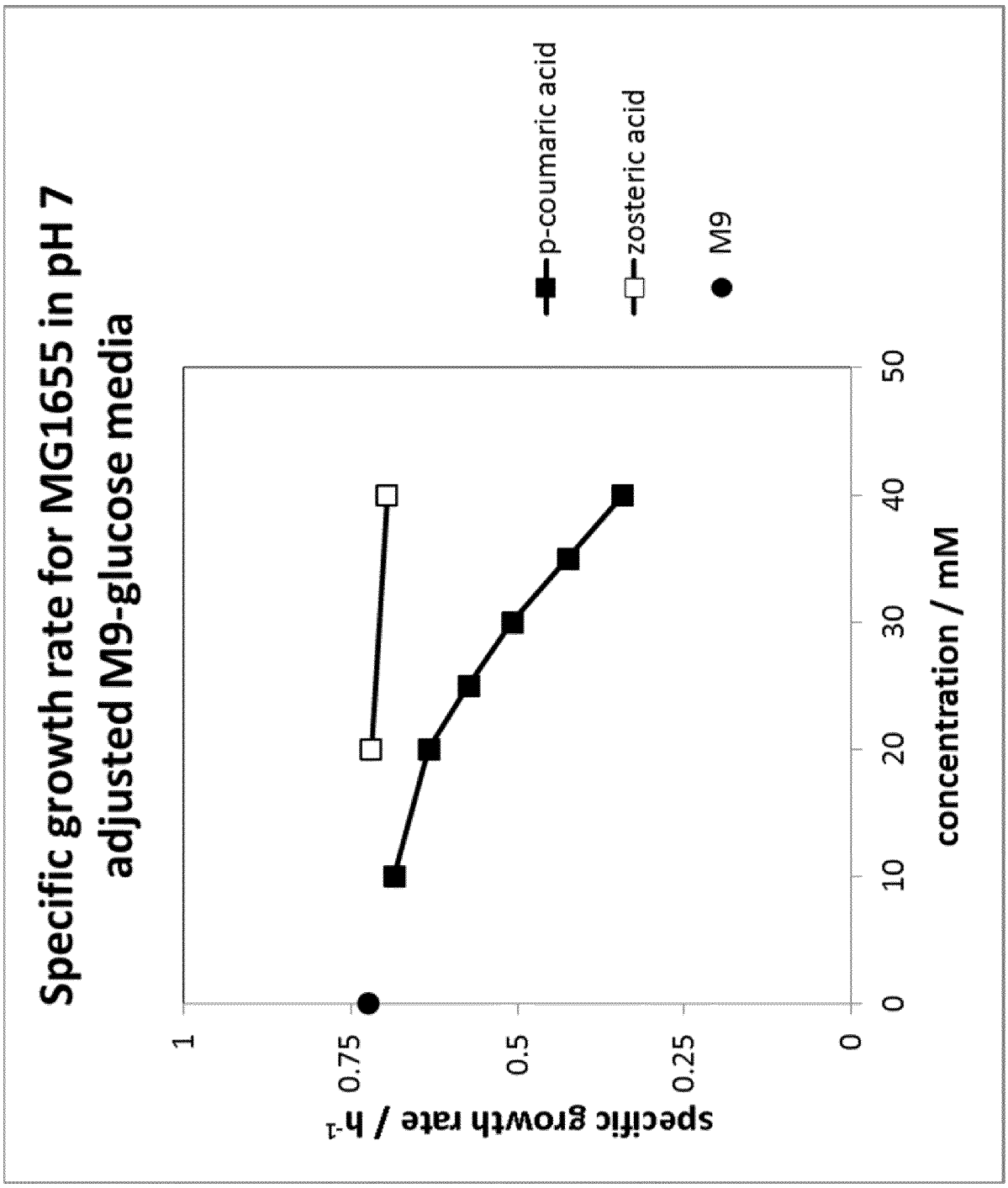

Specific growth rate for MG1655 in pH 7 adjusted M9-glucose media
p-coumaric acid
zosteric acid
M9
specific growth rate / h-1
concentration / mM
0
0.25
0.5
0.75
1
0
10
20
30
40
50

Figure 4

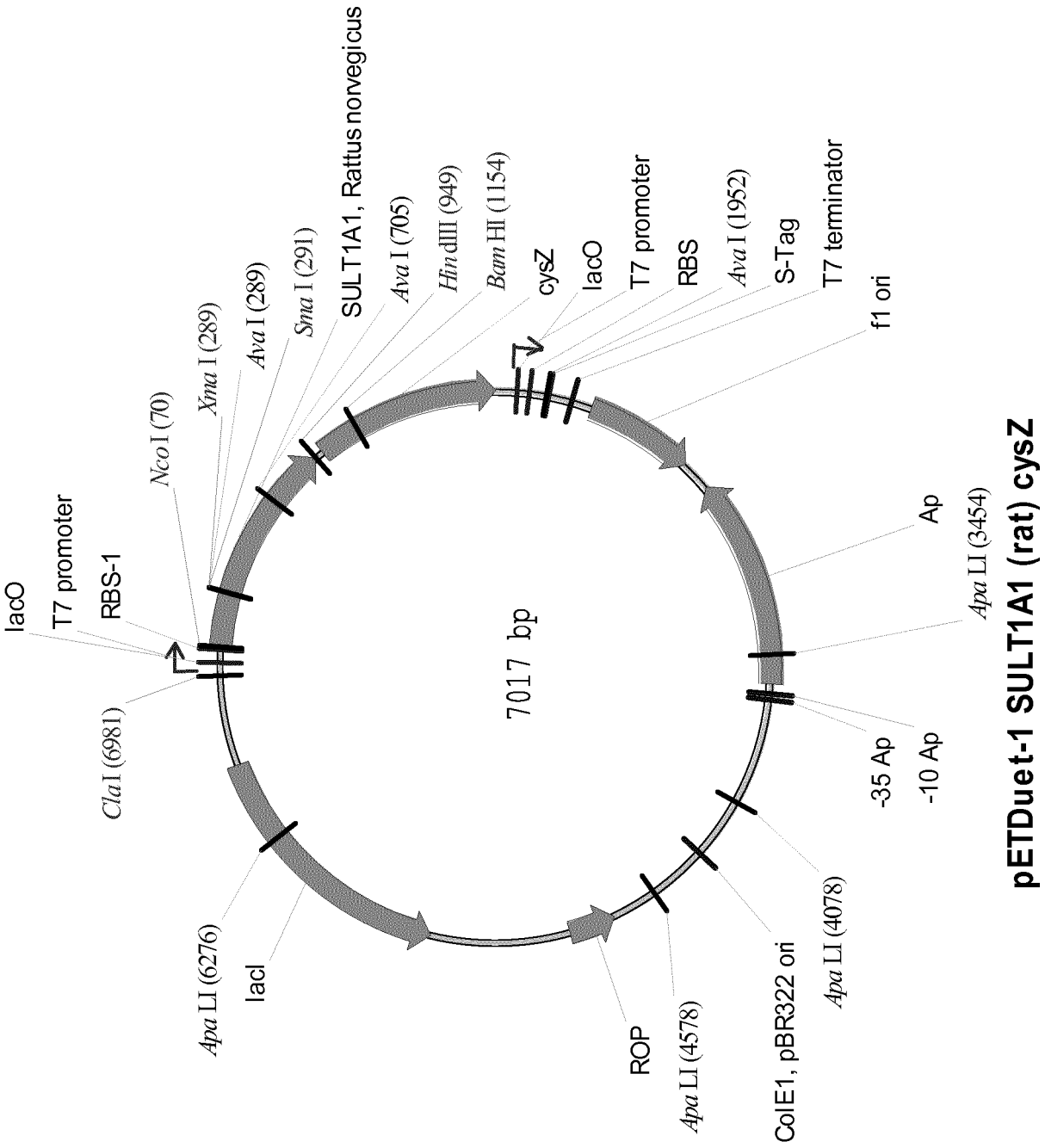
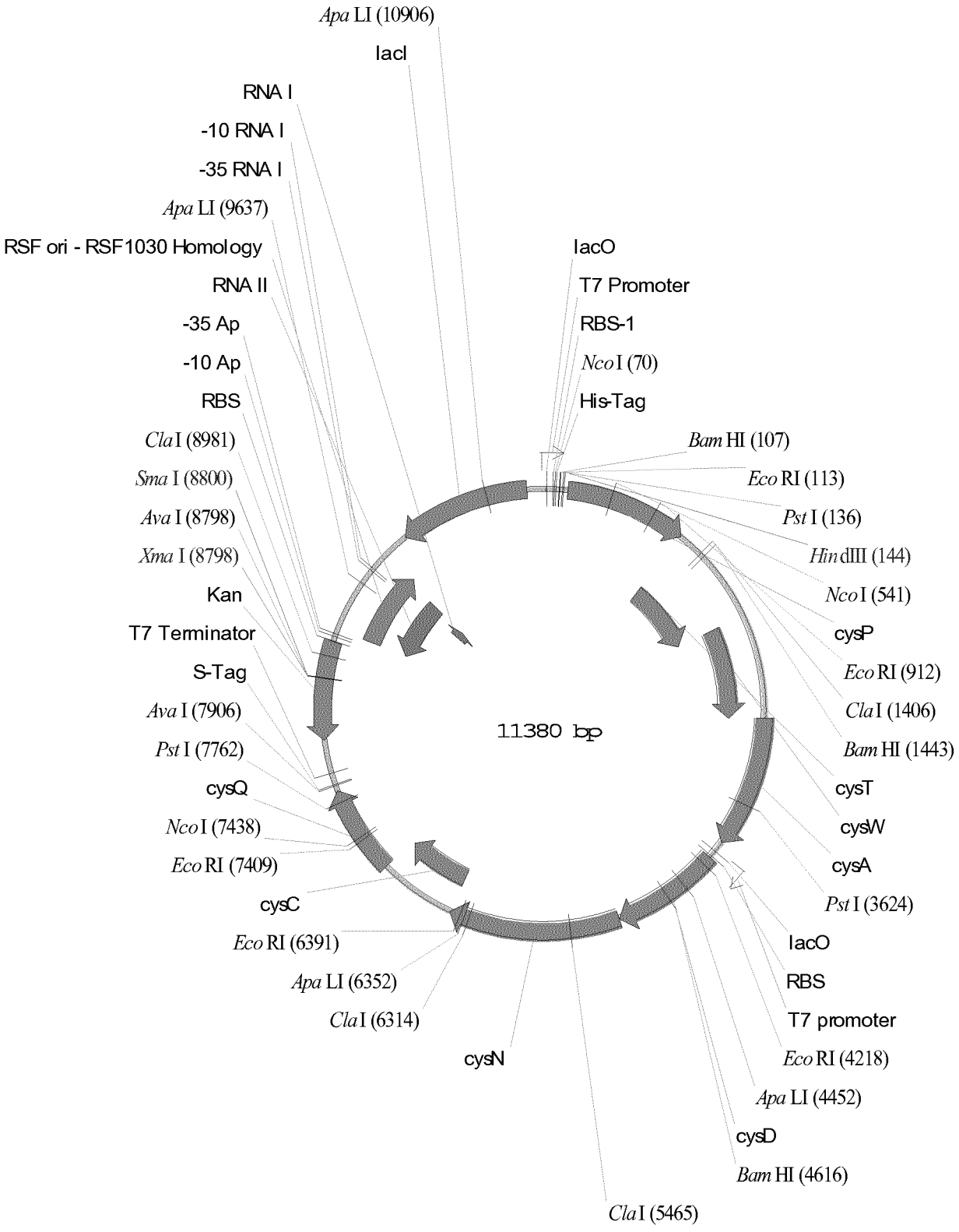

pETDuet-1 SULT1A1 (rat) cysZ
7017 bp
lacO
T7 promoter
RBS-1
Nco I (70)
Xma I (289)
Ava I (289)
Sma I (291)
SULT1A1, Rattus norvegicus
Ava I (705)
Hin dIII (949)
Bam HI (1154)
cysZ
lacO
T7 promoter
RBS
Ava I (1952)
S-Tag
T7 terminator
f1 ori
Ap
Apa LI (3454)
-35 Ap
-10 Ap
Apa LI (4078)
ColE1, pBR322 ori
Apa LI (4578)
ROP
Apa LI (6276)
lacI
Cla I (6981)

Figure 5



pRSFDuet-1 MCS1::cysPTWA MCS2::cysDNCQ

Figure 6

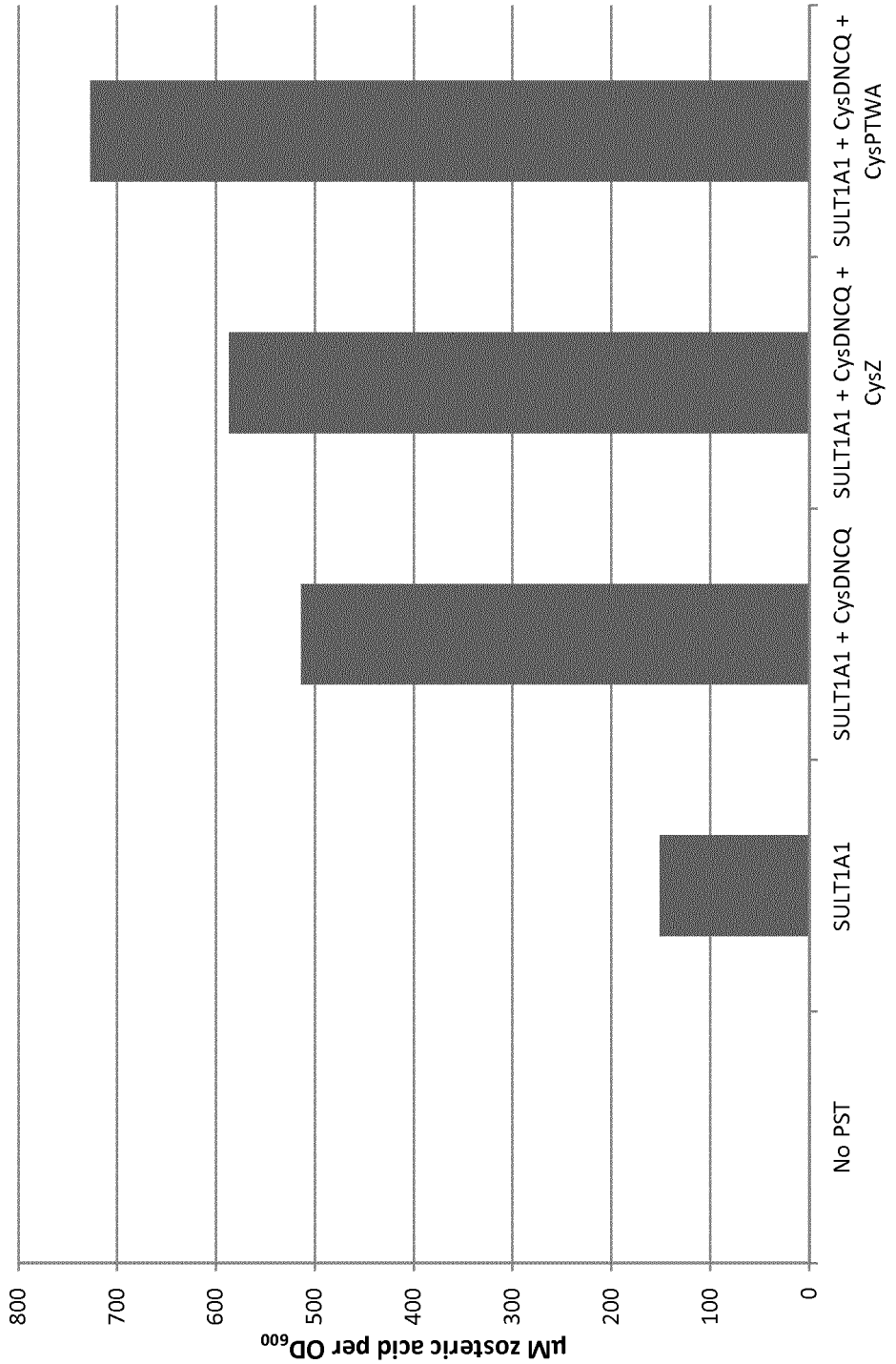

800
700
600
500
400
300
200
100
0
µM zosteric acid per OD600
No PST
SULT1A1
SULT1A1 + CysDNCQ
SULT1A1 + CysDNCQ + CysZ
SULT1A1 + CysDNCQ + CysPTWA

Figure 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/054346

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12P11/00 C12N9/10 C12P5/00
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12P C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 98/03636 A1 (BIO TECH RESOURCES [US]; GRUND ALAN DOUGLAS [US]; MAURINA BRUNKER JULI) 29 January 1998 (1998-01-29) the whole document	1-16
A	WONG C C ET AL: "Inhibition of hydroxycinnamic acid sulfation by flavonoids and their conjugated metabolites", BIOFACTORS, OXFORD UNIVERSITY PRESS, OXFORD, GB, vol. 39, no. 6, 1 November 2013 (2013-11-01), pages 644-651, XP002734418, ISSN: 0951-6433, DOI: 10.1002/BIOF.1127 [retrieved on 2013-08-24] the whole document	1-16

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 April 2017	09/05/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Schneider, Patrick

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YOSHIFUMI KAWAI ET AL: "p-Hydroxycinnamic acid production directly from cellulose using endoglucanase- and tyrosine ammonia lyase-expressing Streptomyces lividans", MICROBIAL CELL FACTORIES, BIOMED CENTRAL, GB, vol. 12, no. 1, 7 May 2013 (2013-05-07), page 45, XP021151039, ISSN: 1475-2859, DOI: 10.1186/1475-2859-12-45 the whole document	1-16
A,P	WO 2016/026976 A1 (UNIV DANMARKS TEKNISKE [DK]) 25 February 2016 (2016-02-25) the whole document	1-16

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2017/054346

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9803636 A1	29-01-1998	AU 4326397 A	10-02-1998
		US 6225100 B1	01-05-2001
		WO 9803636 A1	29-01-1998
WO 2016026976 A1	25-02-2016	NONE	